UNIVERSITEIT GENT

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Economic Integration and Corruption: Resolving measurement and endogeneity problems through state-space modeling

Samuel Standaert

Supervisor: Prof. Dr. Glenn Rayp (Ghent University)

Submitted at Ghent University
to the Faculty of Economics and Business Administration
in fulfillment of the requirements for the degree of Doctor in Economics

UNIVERSITEIT GENT

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Doctoral committee:

Prof. Dr. Marc De Clercq (Ghent University, President)

Prof. Dr. Patrick Van Kenhove (Ghent University, Academic Secretary)

Prof. Dr. Glenn Rayp (Ghent University, Supervisor)

Prof. Dr. Aart Kraay (The World Bank)

Prof. Dr. Scott Baier (Clemson University)

Prof. Dr. Koen Schoors (Ghent University)

Prof. Dr. Dirk Van de gaer (Ghent University)

Submitted at Ghent University

to the Faculty of Economics and Business Administration

in fulfillment of the requirements for the degree of Doctor in Economics

Acknowledgments

Dear *[Insert name here]*,

Thank you so much for

[Please select one or more options]

- ◦ being my doctoral advisor;
- ◦ reading and commenting on my thesis;
- ◦ funding my research;
- ◦ writing papers with me;
- ◦ taking care of all the administration for me;
- ◦ going swimming, running, playing tiles™;
- ◦ jamming with the Animal Spirits;
- ◦ being a friend;
- ◦ putting up with me;
- ◦ making this whole PhD thing extremely enjoyable.

It's been absolutely invaluable.

Samuel Arthur Simon Ephrem Gaston Standaert

Table of Contents

List of Figures

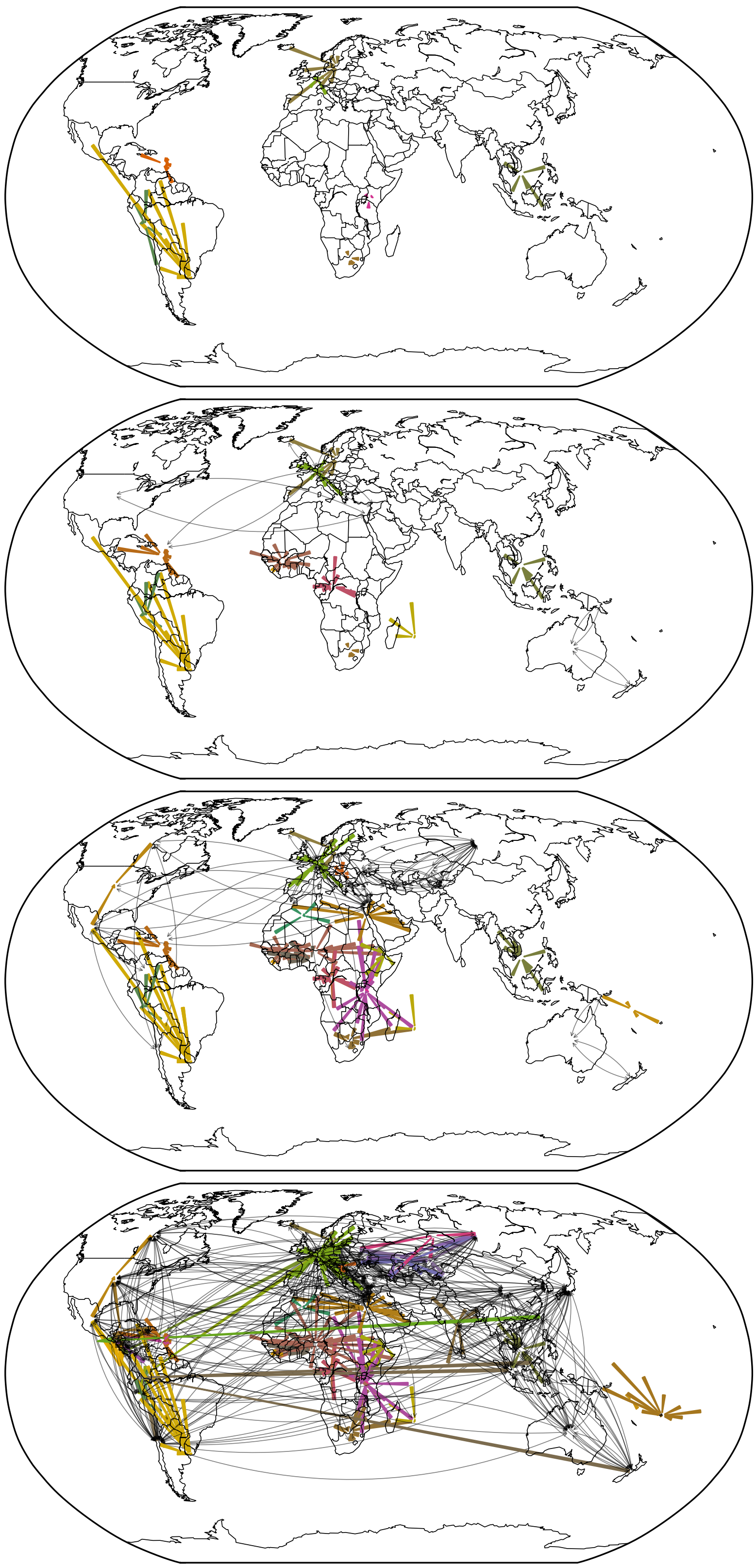
List of Tables

1 | Introduction

1.1 Motivation

Not long after World War Two ended, a small number of European countries started to work together to bring down the barriers to trade between them. While the European Union started out as a very specific cooperation agreement, it grew over a period of 50 years into true economic union covering the bulk of the European continent. During this time, countries all over the world followed suit. This gave rise to a large number of Regional Integration Agreements (RIA) differing in size (the number of countries covered), breadth (the different policy domains covered) and depth (the extent to which each of these domains is harmonized). Throughout this book, the term Regional Integration Agreement will be used to refer to this amalgamated group of intra- and inter-regional free trade agreements, customs unions, common markets and economic unions.

By way of illustration, figure 1.1 plots the proliferation of RIAs over the last 45 years. Bilateral agreements (between two countries) are indicated by black arrows and multilateral agreements (involving more than two partner countries) are represented by connecting those countries to a central node using a colored line. The powerful upswing in the number of agreements over the last 15 years shows that even though the first agreements are more than 60 years old, RIAs are still a present-day tool of modern trade policy. This increase is due in part to the failure of the World Trade Organization's Doha Development round of negotiations to break down barriers to trade. Because of their preferential nature, RIAs make it easier to



(a) 1970

(b) 1985

(c) 2000

(d) 2015

Figure 1.1: The growth of bilateral and multilateral integration agreements over time. Bilateral agreements are indicated by the curved black arrows, while countries that are members of a multilateral ageements are connected to a central node using a colored line.

come to an agreement on a wide range of topics. This enables countries to address issues beyond reductions in tariffs such as removing non-tariff barriers to trade as well as ensure cooperation in other policy domains such as migration, security and environmental policy.

The chapters in this book are bound together by three common threads: they share a common theme, address common methodological issues and/or use similar techniques. After explaining each underlying thread, the remainder of the introduction briefly outlines the contents and contributions of each individual chapter.

The causes and consequences of trade agreements

The overarching theme is an empirical study of the causes and consequences of signing a regional integration agreement.

The theoretical literature on RIAs usually does not discern between causes and consequences since the motives for signing an integration agreement are the expected benefits. Initially, the analysis of the economic effects of trade agreements focused on the static, short term welfare effects caused by the change in relative tariff rates (Viner, 1950). Trade agreements would only positively affect welfare to the extent that they didn't divert too much trade from the rest of the world. However, if the trade agreement's main effect was to benefit the partner countries' companies over those in the rest of the world, an increase in bilateral trade could nevertheless cause a decrease in overall welfare. As a result, many of the ex-ante indicators of the welfare effects of a trade agreement compared the compatibility of both countries' economies with their compatibility with the rest of the world.

The dynamic effects of trade agreements focussed on the small, but cumulative changes a trade agreement could have on growth. By bringing the local markets closer together, RIAs could bring about economies of scale that could lead to a cascade of positive effects on welfare. Even though it pertained African integration, the following quote from Foroutan and Pritchett (1993, p.234) sums up the underlying argument behind dynamic effects: *"Imagine subdividing Belgium into forty-something independent countries, each with its own isolated goods and factor* *markets, a different public administration, currency, language, fiscal and monetary authorities, army, plus a very inefficient inter-country transportation network. Economists would contend that the welfare of individuals would surely be reduced."* Because of their breadth, RIAs are a particularly well-suited tool to start tackling these barriers to trade and as a result carry huge potential welfare benefits.

Finally, the new regionalism theories were concerned with the political process behind the negotiations of RIAs and looked at the effects beyond those on trade. The assumption of a monolithic, welfare-maximizing government was abandoned, allowing for various political motives to drive the formation of RTAs. Examples include using trade agreements to enhance regional cooperation or lock in domestic reform programs (Maggi and Rodríguez-Clare, 1998; Mansfield et al., 2002). Other authors argued that trade agreements could also be used for more nefarious purposes. Think of situations where private firms bribe politicians into closing agreements that are not in the common interest (Grossman and Helpman, 1995) or where politicians themselves use the agreements to strengthen their tenuous domestic political position (Söderbaum, 2004).

Unlike the theoretical literature, the empirical studies of trade agreements can for the most part be separated into studies explaining the formation of trade agreements and those studying its effects. Broadly speaking, the former rely on (ordered) probit models to explain ex post which factors enticed countries to sign trade agreements. An important finding has been that while economic welfare is what matters in the long run, political variables are the deciding factor when it comes to the timing of the agreements (Baier et al., 2007). We link up with this literature in the second chapter where we look at the influence of corruption on the formation of African RIAs.

The main mechanism used to study the effects of trade agreements has been to include various types of RIA dummies into gravity models explaining bilateral trade flows. However, due to various methodological problems the results have been strongly divergent. To date there is no overall agreed upon way of estimating the effects of trade agreements on trade (Baier and Bergstrand, 2009). We address this in chapter 6 where we try to merge the empirical literature on the causes and consequences by estimating the probit model and gravity equation simultaneously.

Methodological issues

The two recurring methodological issues that are centered on in this thesis are simultaneity (or endogeneity) and data quality. Simultaneity is a situation were one of the explanatory variables affects, but is also affected by the to-be-explained variable. For example in the last chapter we look at the effect of integration agreements on trade while taking into account that these agreements are more likely to occur between countries that already trade intensively. Ignoring this source of endogeneity has been found to significantly bias the estimated effect of RIAs downwards (e.g. Magee, 2003; Baier and Bergstrand, 2002, 2004b, 2007, 2009; Egger et al., 2011).
A similar simultaneity problem is addressed in chapter 2. There are a number of theories that find that (under certain conditions) an increase in the level of corruption will raise the likelihood of a RIA. Either because the more corrupt government would try to use the agreements to extract rents (e.g. Grossman and Helpman, 1995) or because the increase in corruption would motivate the government to fight corruption using an integration agreement (e.g. Maggi and Rodríguez-Clare, 1998). In the first case, this would lead to a subsequent increase in corruption, while the opposite would happen in the second case. Regardless of the direction of the effect this simultaneity has to be addressed, especially because not all indicators of the level of corruption predate the signing of the agreement.
The second econometric issue central to this thesis is that of data quality. As Morgenstern (1962) emphasized, many of the variables used in economic analyses are only estimates with a certain and at times significant measurement error. Ravallion (2010, p.2) focuses specifically on the correct construction and use of 'mashup' indices: indexes that combine a set of variables that reflect various dimensions of some unobserved concept, but whose construction does not rely on any theoretical literature. The indicators of corruption perception used in the second chapter are an example of a mashup index. Incautious use of these indicators can lead to the

wrong conclusion, which becomes evident in chapter 2 where the different indicators of corruption can sometimes come to strongly different conclusions.

While often overlooked, the same problem afflicts data on trade flows and indicators of regional integration. Chapters 3 to 5 deal with the creation of a (mashup) index of the level of corruption and integration from a number of imprecise indicators. In the construction of these indicators, we pay specific attention to the issues brought forward by Ravallion (2010): i.e. conceptual clarity, transparency, robustness and policy relevance. We start with a very clear definition of what we are trying to measure and what we aim to use it for. The weight of each component of the indexes is determined statistically, moving any trade-off between different components outside of our direct control. However, to make the interpretation of the index as clear as possible, we subsequently discuss the weight each component was assigned.

The statistical weights notwithstanding, there remain certain modeling choices in each index. For example, do you allow cross-correlation between the measurement errors of the different components? Can the level of persistence in corruption or integration differ in each country? etc. We test the robustness of our results for these choices and employ Bayesian model selection techniques to see which option is best supported by the data. The way in which we deal with missing data also negates the need to impute any of the underlying data series ex ante. Instead, information can simply be entered incompletely and the index will reflect the completeness and reliability of the underlying dataset.

An important question that needs to be resolved is how these indexes can be used in such a way that their underlying uncertainty is fully taken into account. Firstly, we make sure that we do not simply report the values of the index, but each time also give an estimate of its reliability. Moreover, we produce a dataset that contains hundreds of random draws from the distribution of the index such that any subsequent analysis can correctly take its uncertainty into account. In the analyses in this thesis we do this using Bayesian Gibbs sampling, but other possibilities include multiple imputation (cf. Desbordes and Koop, 2014).

In terms of (policy) relevance, one of the main advantages of the indexes is the fact that the significance of changes over time and differences between countries can be statistically determined. This enables for example the creation of 'significant' rankings that reflect only the robust differences between countries (cf. Høyland et al., 2012). This ensures that any change in the ranking of countries carries actual meaning and is distinct from 'noise' in the measurement errors.

The state-space model

The state-space model is the methodological foundation of the estimation models used in chapters 3 to 6. Originally designed to track the flightpath of rockets, the state-space model is essentially a method to compute the distribution of one or more unknown state variables from various (imprecise) measurement variables. It relies on two equations, named the state and measurement equations. Using the original example, the first equation predicts the position of the rocket (the unknown state) using the information on where it was previously. The second equation explains how the various measurements are related to the actual position of the rocket. For example, the measurements from a satellite dish might be more accurate than those of a radar, or might be better at gauging latitude and longitude but score worse at measuring the height, etc.

To get the best estimate of where the rocket is now, the information on the previous and future position should be used. Because these are also unknown, the optimal estimate would require the entire model to be solved at once. However, Rudolf Kalman (1960) showed that under certain assumptions the state-space could be solved iteratively. Iterating forward, the Kalman filter would update the estimate of the state variable using all information available in the past. The Kalman smoother would then iterate backwards to update these estimates using all information available in the entire sample (Kim and Nelson, 1999; Durbin and Koopman, 2012).

This way of solving a complex dynamic system iteratively can be used on many problems, provided that they can be written as a state-space model. These range from composing an index of corruption or regional integration (chapters 3 to 5) to estimating the willingness to sign trade agreements in chapter 6. Additional advantages are that any subsequent computation or estimation can take the uncertainty of this estimate into account. Moreover, the state-space model has an elegant solution to the problem of missing data points that does not require any imputations or other manipulations to the data.

1.2 Contributions and results

The chapters in this book are listed in chronological order. On the one hand, organizing them this way shows how the original research question evolved over time. It also explains why for example the new corruption index of chapter 3 is not used when analyzing the link between corruption and RIAs in the preceding chapter.

The research question analyzed in the **second chapter** stems from the apparent contradiction in intra-African integration agreements. Even though all African countries are a member of one or more multilateral RIAs, intra-African trade is surprisingly low. In most cases, these agreements failed to produce a positive effect on trade or welfare, which is what most ex ante economic indicators also predicted.

The failure of the more traditional economic motives to explain African RIAs opened up the opportunity to test more political motives identified in the theory, specifically the importance of rent-seeking behavior. Interestingly, the theories on rent-seeking and RIAs find that the level of corruption is posited to have a positive effect on the willingness to enter into RIAs, regardless of whether it was meant to foster corruption or prevent it. In addition, most theoretical models found a strong non-monotonic effect of corruption. At intermediate values an increase in corruption would raise the probability of a RIA, but if corruption became too high this effect would be inverted.

In our study we found that corruption has a significant effect on the willingness to join trade agreements. While small, the effect of corruption outperformed most economic variables such as the difference in the capital-labor ratio or GDPs. This finding was robust for different indicators of corruption, correcting for endogeneity and the fact that most African trade agreements involve more than two partner countries. That being said, we failed to observe the non-monotonic effect predicted by some of the theoretical models and the economic relevance of corruption was small. Distance and other geographical factors remained the strongest explanatory forces of African integration.

Except for a few indicators in favor of the model of Grossman and Helpman (1995), the analysis of African RIAs was not able to discern between the different rent-seeking motives of RIAs. The natural follow-up question would be to find out under which circumstances trade agreements fostered or hindered corruption. However, it became clear that the indicators of corruption used in chapter 2 would not suffice. The Worldwide Governance Indicators index of corruption (WGI), while strong methodologically, was only available from 2002 onwards on a yearly basis. Transparency International's Corruption Perceptions Index (CPI) started in 1995 but had serious selection bias issues and could not be used to make comparisons over time. The International Country Risk Guide's index of corruption on the other hand started in 1984, but covered fewer countries. In addition, ICRG came from a single source, while CPI and WGI combined many different indicators of corruption.

One of the principle problems when combining different indicators of corruption is that of missing observations. The group of available indicators differs in each year as do the countries covered by each indicator. Ignoring this in the construction of the index makes it impossible to know whether a change in the index is caused by a change in corruption or by differences in the availability. The CPI solved this problem at the cost of losing the ability to compare the index over time as well as introducing a strong selection bias in each individual year. WGI on the other hand combined information of several years, lowering the frequency of the index in the initial years.

Instead, in **chapter 3** we turned to a source of information that was not being used in the creation other indexes: the strong time-dependence in the level of corruption. Adding this time-dependence to the methodology of the WGI resulted in a state-space model. To estimate it, we used a Bayesian Gibbs sampling algorithm that allowed us to split up a complex probability into several much easier to solve subcomponents. The new Bayesian Corruption Indicator was able to significantly increase coverage, while producing more stable estimates with smaller confidence intervals.

The methodology of the state-space model proved applicable to a wider range of problems. Specifically, in the **fourth chapter** we argue that it can be used to measure the level of economic integration. The state-space model weighs each indicator's influence statistically, allowing us capture the overall evolution of such a multidimensional concept as integration. Furthermore, the model can easily be extended to combine for example dichotomous with continuous indicators without having to impose strong assumptions. This would for example allow the addition of indicators of institutional integration in a straightforward and objective manner.

As an example, we used the state-space model to build an index of Actual Economic Integration that captures integration in terms of trade in goods and services, migration and financial flows. Starting in the mid 1980s, the index looks at integration from the perspective of OECD member countries. Using this index in a weighted directed network, Germany was identified as the most central country in the OECD network with the USA as a close second. It also shows the gradual rise in importance of two non-OECD members during the last two decades, namely China and Russia. Using this index as the dependent variable in a structural gravity model revealed that the European Union (EU) and the North American Free Trade Agreement (Nafta) were closed between countries that were already highly integrated. Nevertheless, these agreements succeeded in further raising the level of economic integration both in the short and long term.

The **fifth chapter** limited the index to measure only trade integration, which allowed us to expand analysis to the entire world, starting in the 1880s. The state-space model's solution to incomplete information played a key role in extending the coverage before 1950 compared to other indicators of trade integration. For example, the number of country-pairs covered in each year before 1950 was quadrupled relative to the Head-Ries index. The increased coverage enabled the comparison of the change in the importance of distance during the pre- and post-World War globalization waves.

The concept of geographic neutrality predicts that the influence of distance on the global trade pattern would diminish as the world is becoming more globalized. However, over the last three decades a wide range of studies has found an increase in the importance of distance during the second half of the twentieth century (*the distance puzzle*). In line with the theories of O'Rourke (2009) and Jacks et al. (2011), we found that this increase is not caused by methodological issues, but is an actual feature of the changing world trade network. The first globalization wave (1880-1914) was driven by decreasing trade costs, significantly lowering the importance of distance over time. In contrast, the second wave (during the second half of the 20^{th} century) was induced by increased production capacity and geopolitical changes centered on Western Europe and North America. While we confirm the increase in the effect of distance on trade during this period, we found that it is dominated by the strong decrease during the first globalization wave.

Finally, **the last chapter** returns to the analysis of integration agreements, specifically we look for a way in which the simultaneity of trade and trade agreements can be modeled. Initial attempts to use instrumental variables proved unreliable, causing many to look for alternative solutions to address this issue. In the last chapter, we bring together the gravity equations explaining trade and the probit regressions explaining RIAs by modeling them simultaneously.

Estimation once again relied on Gibbs sampling techniques, where in a first step a state-space model was used to generate the willingness to sign a trade agreement. The latter was subsequently modeled endogenously with trade and its parameters are estimated in a vector autoregression model (VAR). This qualitative VAR model allowed us to capture the interdependence of trade and trade agreements without having to resort to instrumental variables. Moreover, it enabled us to take the endogenous nature of other control variables like GDPs and the capital labor ratio into account. Our preliminary findings confirmed the simultaneity of trade and trade agreements: an increase in trade raised the probability of an agreement and vice versa, although the response differed over specific continents. Both variables displayed strong dynamic behavior and overall the parameters on the control variables followed expectations.

Using the estimated parameters, it is possible to calculate what trade and GDP would have been if no agreement had been signed. By comparing the actual flows with these counterfactuals, we can compute the average treatment effect of signing a trade agreement. Preliminary results indicate that the effects are relatively small compared to what is typically found in the literature. Trade increases with 10% after one year and 40% after 5 years, after which its rate of growth drops off. GDP initially does not seem to be strongly affected by the agreement. After 35 years, trade flows have increased with about 80% while GDP increased with only 10%. However, these findings are very preliminary. Before estimates can be stated with any confidence, the theoretical foundations of the estimation model should be more firmly secured.

References

Baier, S.L., Bergstrand, J.H. (2002) On the endogeneity of international trade flows and free trade agreements.

Baier, S.L. and Bergstrand, J.H. (2004) Do free trade agreements actually increases members' international trade?

Baier, S.L. and Bergstrand, J.H. (2007) Do free trade agreements actually increases members' international trade? *Journal of International Economics* 71:72–95.

Baier, S.L., Bergstrand, J.H. (2009) Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of International Economics* 77:63–76.

Baier, S.L., Bergstrand, J.H. and Egger, P. (2007) The new regionalism: causes and consequences. *Economie Internationale* 100:9–29.

Durbin, J. and Koopman, S. (2012) *Time Series Analysis by State Space Methods, 2nd edition.* Oxford University Press, Oxford.

Desbordes, R. and Koop, G. (2014) The known unknowns of governance. Stratchclyde discussion papers in economics 14-07.

Egger, P., Larch, M., Staub, K.E., and Winkelmann, R. (2011) The trade effects of endogenous preferential trade agreements. *American economic Journal: Economic Policy* 113–143.

Foroutan, F. (1993) Intra-Sub-Saharan African Trade: is it too Little? In de Melo, J. and Panagariya, A. (editors) *New Dimensions in Regional Integration,* Cambridge University Press, chapter 8, p. 234–311.

Grossman, G.M., Helpman, E. (1995) The politics of free-trade agreements. *The American Economic Review* 85(4):667–23.

Jacks, D.S., Meissner, C.M. and Novy, D. (2011) Trade booms, trade busts, and trade costs. *Journal of International Economics* 83(2):185–201.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems *Transactions ASME Journal of Basic Engineering* D82: 35–45.

Kim, C.J. and Nelson, C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* MIT Press, Cambrigde.

Mansfield, E.D., Milner, H.V. and Rosendorff, B.P. (2002) Why democracies cooperate more: Electoral control and international trade agreements. *International Organization* 56(3):477–513.

Magee, C. S. (2003) Endogenous Preferential Trade Agreements: An Empirical Analysis. *Contributions to Economic Analysis & Policy* 2(1):1166–1217.

Maggi, G. and Rodríguez-Clare, A. (1998) The value of trade agreements in the presence of political pressures. *Journal of Political Economy* 106(3): 574–601.

Morgenstern, O (1963) *On the accuracy of economic observations.* Princeton University Press, New Yersey.

O'Rourke, K. (2009) Politics and trade: lessons from past globalisations. Bruegel.

Ravallion, M. (2010) Mashup indices of Development. World Bank Policy Research Paper 5432.

Söderbaum, F. (2004) *The Political Economy of Regionalism: The Case of Southern Africa.* Palgrave Macmillan, New York.

Viner, J. (1950) *The Customs Union Issue.* Stevens and Sons, London.

2 | Multilateral trade agreements in Africa - Exploring the role of rent-seeking behavior[1]

Abstract

This chapter explores the motives behind the formation of intra-African regional integration agreements (RIAs). We focus specifically on rents because they can explain the drive for integration even in the absence of a positive effect on welfare. Whether they are meant to foster rent-seeking or to suppress it, the level of corruption is posited to have a positive effect on the willingness to enter into RIAs. However, previous empirical studies have come to contradictory conclusions. We find that corruption has a significant effect on the willingness to join trade agreements. While small, the effect of corruption outperforms most economic variables. Nevertheless, distance and other geographical factors remain the strongest explanatory forces of African integration.
Keywords: African regional integration; Rent-seeking; Multilateral trade agreements; Corruption.

JEL classification: Q27; D73; F13; F54.

[1]This chapter is the results of joint work together with Prof. Dr. Glenn Rayp.

2.1 Introduction

Regional integration has been very popular in Africa over the last 50 years. Every country is part of at least one regional integration agreement (RIA) and on average an African country is member of four agreements (The World Bank, 2005). Yet, it is hard to reconcile this enthusiasm for regional integration with its results. Most indicators show that African economies are barely integrated. Tariff reduction schemes are backlogged, rules of origin are extremely restrictive and cross-border transportation facilities are either inadequate, or missing altogether. As a result, the level of intra-regional trade of most African RIAs rarely exceeds 10% (relative to around half of all trade in NAFTA or the EU-27), and in some cases it even fell after signing the agreements (UNU-CRIS, 2006).

Theoretically, the reasons for African integration have never been very compelling to begin with. First of all, most African countries do not produce any of the products that are of interest to neighboring countries. The bulk of African trade is with developed economies, in particular the European Union. The African trade patterns are not complementary and static analysis warns that this will most likely result in trade diversion and hence lower welfare. Similarly, the long term dynamic effects are unlikely to be strong and many of the new regionalism theories are conditional on strong economic integration.

Rent-seeking behavior on the other hand can provide a valid alternative explanation for the interest in African integration in the form of the rent-seizing, rent-shielding and rent-destruction hypotheses. They can account for the strong interest in integration in the absence of a positive effect on welfare, as well as for the lack of progress in breaking down the barriers to trade. The rent-seizure hypothesis states that RIAs are used to set up rent-extracting mechanisms. The agreements bestow extensive powers on the negotiating parties, which combined with the absence of an increase in welfare creates an ideal environment for lobbying and in the worst case bribery. Conversely, rent-shielding remonstrates that the government can use RIAs to protect parts of the economy by removing them from the direct control of the government. Similarly, rent-destruction claims that RIAs increase competition which curtails rents. Depending on how benevolent the government is, all three predict a positive effect of corruption on the formation of trade agreements.

While there exist ample examples of RIAs being misused for rent-seizure purposes, the evidence remains largely anecdotal. A couple of studies have regressed the desire to enter into a free trade agreement on the level of corruption. Wu (2006) found that an increase in the level of corruption increases the probability of joining a FTA in that year, but ignored characteristics of the partner country. In contrast, Endoh (2006) found that the quality of governance (including the absence of corruption) and the probability of an agreement are positively related. A possible reason for the contradictory results is the fact that these models only looked at a monotonic relation. In contrast, most theoretical models predict that rent-seeking has a non-linear effect and may interact with relative GDP or other characteristics. Arcand et al. (2011) found that government's welfare mindedness has a myriad of significant interaction terms, among which its squared value and the bargaining power of the government.

The goal of this chapter is to see whether the level of corruption can be identified as a statistically significant factor in the decision to enter a free trade agreement in Africa. In doing so, we also build on papers by Mansfield et al. (2002), Baier and Bergstrand (2004a) and Márquez-Ramos et al. (2011) that try to determine the reasons behind the formation of regional integration agreements.

The approach followed in this chapter differs from that employed in the papers listed above in a number of ways. Firstly, it is centered on intra-African integration and as we will show, there is significant regional heterogeneity in the reasons for entering free trade agreements. The failure of the traditional economic motivations in explaining African RIAs opens up opportunities to test more political motives. Secondly, we explicitly take into account that the intra-African trade agreements involve more than two partner countries. Moreover, many of these agreements are overlapping, meaning that a country can be in two trade agreements with partner countries that have not signed an agreement themselves. To this end, the bilateral estimation framework is adjusted to deal with the heteroscedasticity caused by the non-nested multilateral trade agreements. Thirdly, we do not infer the level of rent-seeking from tariff rates like Arcand et al. (2011), but instead use different indicators of corruption perception as a proxy for the level of rent-seeking. Finally, we address the endogeneity of the relationship between corruption and trade integration in a number of ways: using lagged values, a control function approach as well as an instrumental variable probit regression model.

The remainder of this chapter is organized as follows. Section two gives an overview of the different reasons for economic integration and their merits and demerits in explaining African integration. Next, we expand on the idea of rent-seizure, rent-shielding and rent-destruction as driving forces of integration. Sections four and five cover the data and econometric method used, after which we discuss our findings.

2.2 Irrational exuberance?

2.2.1 Static and Dynamic effects

In his seminal work, Viner (1950) states that lowering the barriers to trade on an imported good for a specific group of partner countries will have two conflicting effects on domestic welfare: trade creation which raises welfare and trade diversion which lowers it. Whether a RIA will raise or lower welfare depends on the relative size of these two effects over all sectors covered by the RIA.

The problem with African integration is that intra-African trade often fell after the agreements were closed (Iapadre and Luchetti, 2010; Yeats, 1998). With the exception of South Africa and a few other more developed nations, exports are focused on primary goods, while imports are for the most part manufactured goods. The non-complementarity in African trade was high and even under the most favorable assumptions not likely to change rapidly (Yeats, 1998). Export infrastructure was aimed at the developed world, and local trade infrastructure either missing or inadequate, suggesting the opportunity costs of intra-African RIAs was relatively high

(Yang and Gupta, 2007). Carrere (2004) found that, when controlling for other factors that adversely affect bilateral trade (e.g. inadequate infrastructure, low GDP, etc.), the trade agreements did promote intra-African trade. However, in free trade agreements this effect was mostly driven by trade diversion, as opposed to customs unions where trade creation prevailed.

The argument of long-term economies of scale (or dynamic effects) also has some flaws. First of all, even if we were to unite all sub-Saharan markets, the combined GDP would still be small, especially given the size of the African continent. For instance, in 2009 the combined GDP of all sub-Saharan countries roughly equaled that of the state of New York. Secondly, to fully integrate a slew of problems would have to be conquered: different languages, currencies, rules and regulations, practically non-existent transnational transportation facilities, etc. Circumventing or breaking down these barriers to trade is extremely expensive in time, money and human capital (Foroutan and Pritchett, 1993). In short, the cost of attaining the level of integration that is needed to produce economies of scale outweighs its benefits in the short and medium-long term.

To exacerbate the problem, the distribution of the benefits of bilateral trade liberalization is highly uneven. Most agreements are dominated by a *hegemon*: a country whose GDP is the multiple of that of other members, that is more industrialized, has higher tariff rates and often is the sole producer (Carrere, 2004). As a result, trade diversion will be high with most benefits accruing to the hegemon, leaving the smaller partner countries to pay for its increase in welfare. Furthermore, the location theory of Krugman and Venables (1989) predicts that removing barriers to trade in this setting will lead firms to relocate to the biggest market, especially when those barriers are taken down gradually. Combining distributional effects with the limited impact on trade and growth means that regional integration in Africa becomes a near zero-sum game. That these distributional problems are not without consequences, was all too clear in the East African Community (EAC) where they led to its dismantlement in 1977.

In order to exemplify the problem standard economic theory has in explaining

African trade agreements, columns one to five of table 2.1 replicate the regressions in Baier and Bergstrand (2004a, table 1) for the African continent. Only the distance-related characteristics (natural and remoteness) have the expected signs and are significant. While significant the effect of the sum of, and difference in GDP runs opposite to what Baier and Bergstrand find. DKL and DROWKL are the difference in capital ratios between the two countries and the difference with the rest of the world, respectively (cf. infra). They are insignificant and the likelihood ratio test rejects the joint significance of these non-distance related variables, confirming the regional heterogeneity in explaining trade integration.

2.2.2 New regionalism theories

The new regionalism theories list other possible reasons for integration that go beyond advantages to trade. A first one is that RIAs can help bring peace to the war-struck African continent. Undoubtedly inspired by the European successes, this argument nevertheless falls somewhat short. First of all, the pacifying effect of the European Union is built on the fact that its member countries have strong economic ties, which would make war *"not only unthinkable, but materially impossible"* (Schuman, 1950). However, as the previous section showed, this economic deterrent is largely missing in African integration. Moreover, in sub-Saharan Africa domestic forces are often a much bigger threat to political stability than international ones. Historically speaking, the probability of revolutions, coups and civil wars is a far greater than an all-out war between nations; A fact illustrated by the recent splitting up of the Republic of the Sudan (Söderbaum, 2004).

Another often cited effect of regional integration is that it can strengthen policy credibility and help lock in domestic reform (Maggi and Rodríguez-Clare, 1998; Mansfield et al., 2002). By being self-enforcing or allowing compensating action or punishment, a RIA can raise the perceived legitimacy of domestic reform programs. However, finding a strong enough enforcement mechanism is not an easy task. Case in point is the serious backlog in most reform programs and the increasing use

Table 2.1: Economic and political explanations of RIAs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Natural	0.485***	0.554***	0.554***	0.555***	0.554***	0.513***	0.496***
	(16.16)	(16.05)	(16.03)	(16.06)	(16.03)	(14.13)	(13.36)
Remote	1.057***	1.220***	1.241***	1.193***	1.241***	1.147***	1.275***
	(7.79)	(7.70)	(7.81)	(7.41)	(7.53)	(6.71)	(7.05)
Adjacent	–	–	–	–	–	0.316***	0.297***
						(3.68)	(3.22)
Landlocked	–	–	–	–	–	−0.023	$-2.02e-4$
						(−0.63)	(−0.01)
GDP_{av}	–	−0.0443*	–	−0.0443*	–	−0.0626**	–
		(-1.75)		(−1.75)		(−2.40)	
GDP_{diff}	–	0.0388**	–	0.0384**	–	0.0490**	–
		(2.08)		(2.06)		(2.56)	
GDP_a	–	–	$6.89e-3$	–	$6.08e-3$	–	$6.83e-3$
			(0.46)		(0.41)		(0.40)
GDP_b	–	–	$-7.75e-3$	–	$7.13e-3$	–	$6.47e-3$
			(−0.56)		(−0.51)		(0.40)
DKL	–	–	–	0.0493	0.048	$9.01e-3$	0.021
				(0.78)	(0.76)	(0.44)	(0.97)
DKL^2	–	–	–	−0.0147	−0.0145	–	–
				(−0.71)	(−0.71)		
DROWKL	–	–	–	−0.015	−0.015	−0.014	−0.020
				(−1.01)	(−1.03)	(−0.88)	(−1.27)
Colony	–	–	–	–	–	0.144***	0.131***
						(3.94)	(3.49)
$Polity_{av}$	–	–	–	–	–	$-7.08e-3$	–
						(−1.30)	
$Polity_a$	–	–	–	–	–	–	$9.86e-3$**
							(−2.22)
$Polity_b$	–	–	–	–	–	–	$3.81e-4$
							(0.10)
Observations	1378	1128	1128	1128	1128	1127	1035
Constant	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
log likelihood	−783.7	−602.5	−604.6	−601.7	−603.8	−587.5	−544.3
LR[a] econ/pol		0.107	0.832	0.305	0.857	$7.79e-4$	$252e-3$

Probit regression on RIAbi. More information on the dependent and explanatory variables can be found in section 2.4. Marginal effects reported with t-statistic between brackets. [a]p-value of the likelihood ratio test of the joint significance of the economic and political variables, i.e. everything except the geographical variables. *, **, *** indicates significance at the 10%, 5% and 1% level.

of (informal) non-tariff barriers aimed at protecting against regional competition (Khandelwal, 2004). Credibility is not an exogenous characteristic, but depends on the economic success of the arrangement (Winters, 2001).

Regional integration and cooperation naturally reinforce each other. However, in the short term, there is a trade-off between cooperation and integration. As the negotiations of RIAs require time, effort, and expertise, they crowd-out the negotiations of other possible cooperation schemes. Additionally, if for example issues of a fair distribution of the profits of the RIA impair the trust between the partner countries, this could threaten cooperation in the long run (e.g. the dismantlement of the EAC).

To assess the explanatory power of some of the new regionalism theories, the last two columns of table 2.1 add a number of political variables also used in Mansfield et al. (2002). In line with what they find, having been colonized by the same colonial power significantly increases the probability of being a member of the same trade agreement. The effect of the level of democracy on the other hand is reversed. While Mansfield et al. find that an increase in the level of democracy has a positive effect on the formation of trade agreements, we find a negative coefficient. However, the coefficients are very small and with the exception of the level of democracy of country a, they are not significant. The addition of the political variables renders the non-distance related variables jointly significant, but this is mostly driven by the addition of one variable: the dummy capturing shared colonial history.

2.3 Rents and RIAs

The difficulties traditional economic motivations have in explaining African RIAs opens up the possibility to test more political motivations that do not necessarily depend on strong economic advantages. Specifically, this chapter focuses on the effect of rent-seeking behavior. The literature proposes three ways in which this could lead to an increase in the number of trade agreements: rent-seizure, rent-destruction and rent-shielding.

2.3.1 Rent-seizure

The rent-seizure hypothesis states that RIAs are used to appropriate rents. They have a big impact on a substantial part of the economy (e.g. influencing market structure, conditions that have to be met to import or export goods, etc.) and the negotiations bestow extensive powers to politicians and bureaucrats. Furthermore, the complexity of the agreements allows corrupt officials to find exploitable loopholes that can easily hide their actions. This combined with the near zero-sum-game outcome of integration creates an ideal breeding ground for political lobbying, corruption and bribery.

Take for example the Economic Community of West African States (ECOWAS), where the regional cooperation tax had been set up to compensate countries for the loss of tariff revenues on intra-regional trade. The system set up turned out to be highly discriminatory and led to fraudulent behavior because the compensation computations were based on highly unreliable data (M'Bet, 1997).

A similar situation occurred with the 'single tax' (*tax unique*) in the Central African Monetary and Customs Union (UDEAC). The official goal of this tax was to foster and protect intra-regional production by limiting domestic and import taxes relative to extra-regional goods. Selected goods from membership countries were taxed once when crossing the border at their respective single tax rate and would then be exempt from all other indirect taxes and import duties. This set-up resulted in an extremely discriminatory system, where not only each firm, but also each good within a firm could be subject to its own tax rate. As a result, the setting of each tax rate was subject to numerous strategic considerations and had little to do with economic considerations (Decaluwe et al., 1997). Even in the case were all tariffs on the goods of the partner country are removed, there are still opportunities for corruption: e.g. officials could be bribed to treat goods from the rest of the world as coming from the partner country.

Regional institutions can also be misused in more direct ways. Decaluwe et al. (1997) provide the example of UDEAC where participants in missions and reunions were so well compensated that civil servants in the member states started submitting dossiers on any pretext to guarantee their attendance at these gatherings. Agendas would be littered with items that allowed the so called experts to attend the head-of-state summits, even though those studies were often of poor quality. Besides wasting money that could otherwise have been spent on more productive goals, this also severely impeded with the workings of the UDEAC institutions.

A number of authors have modeled the government's decision to enter into a free trade agreement, where this choice is determined in part by the effect on overall welfare but also by contributions from lobby groups of the affected industries. In one of the most widely cited papers on this subject, Grossman and Helpman (1995) found that a trade agreement could be possible even in the absence of a positive effect on welfare. Moreover, in this setting the likelihood of an agreement rises as the preference for rents (as opposed to overall welfare) increases, but this is conditional on the relative size of the country.[2]

In the model of Grossman and Helpman (1995) the industry's lobby groups directly offer campaign contributions to those political parties negotiating the trade agreements. This differs from the examples provided where the trade agreements themselves could be used by the government to extract rents. Assuming that these systems were consciously constructed in this way, either other members of the government offered incentives to the negotiators, or the negotiators themselves were setting up systems in which they, or people they appoint, can benefit.

Söderbaum's (2004) regime boosting hypothesis can be seen as a special case of the rent-seizure hypothesis aimed at international rents, rather than domestic rents. It states that governments in a tenuous political position will use regionalism to increase their domestic support. By attending regional summits, signing protocols, etc., the government seeks recognition of its legitimacy abroad, which it then uses to attract foreign aid. Söderbaum gives the example of SADC where national projects were often disguised as regional ones and got funded with donor money.

2.3.2 Rent-destruction

Ornelas (2005) expanded the model of Grossman and Helpman (1995) by arguing that RIAs lead to more competition between countries, reducing the returns to high external tariffs for the import-competing industries. Because lobbies take this into

[2]The possibity of a RIA increased as the weight of overall welfare in the government's objective function decreased relative to that of political contributions. However, this only took effect in the case of the larger of the two countries, as the effect of a trade agreement on welfare was positive for the smaller one (Grossman and Helpman, 1995).

account when deciding whether or not to support a trade agreement, the viability of welfare-reducing free trade agreements is severely impaired. He finds that the higher the government's preference for rents is, the more rents will be destroyed by the RIA and the stronger this rent-destruction effect will play. As a result, welfare-reducing RIAs are only possible at a small subset of intermediate levels of corruption. If social welfare was the only thing of importance, the government would never consider welfare-reducing free trade agreements. With high preference for rents the rent-destruction effect dominates and lobbies would not support the trade agreement. Only at intermediate levels of corruption can the rent-seizure motive prevail over the rent-destruction effect. In other words, when both rent-seizure and rent-destruction effects are in play, the effect of corruption on the probability of an agreement follows an inverted-U shape.

This model was further augmented by Arcand et al. (2011) who also took the strength of the bargaining position of the government into account. Their estimations confirmed the theoretical prediction that the government's preference for rents significantly affect its decision to enter into a trade agreement. Moreover, they found strong non-linearities: welfare mindedness and its square value interact both with the government's bargaining position as well as the relative GDPs, creating a complex network of interaction terms. They measured the government's welfare mindedness and its bargaining position by substituting real tariff data in a model of tariff-setting and solving for the relevant parameters.

Endoh (2006) also worked out a theoretical model detailing the effect of changes in the quality of governance on the formation of RIAs. While his model is based on Grossman and Helpman (1995), the effect of a change in governance is reversed: better governance raises the probability of closing a RIA. The main reason for this is that he uses a different government objective function where import tariffs are treated the same as contributions for lobbyists.[3] The inversion of the effect of rent-seizure illustrates the *cognitive dissonance problem* in these particular polit-

[3]The failure to tax is seen as a measure of weak governance, and is doubly included in the government's objective function: once in the overall welfare and once added to the political contributions.

ical support models: tariff revenues play a crucial role in the mechanisms driving the model even though they are not found to matter that much in the actual decisions of governments of developed countries (Ethier, 2011).

In his empirical analysis, Endoh (2006) composes an index of the quality of governance by taking the average of all Worldwide Governance Indicators, including the level of corruption. In contrast to Arcand et al. (2011) he finds that an improvement in governance raises the probability of an agreement. Wu (2006) argues that an increase in uncertainty will raise the willingness to join a RIA and uses the Corruption Perception Index as one of the proxies for political uncertainty. Like Arcand et al. she finds a significant negative relation. A possible reason for these contradictory findings is that Wu and Endoh do not take into account the inverted-U relation between corruption and RIAs. Alternatively, it could simply be that good governance as a whole has a different effect than corruption on its own. The regressions in this chapter will allow for a non-monotonic relation between corruption and RIAs. However, unlike Arcand et al. we will use indicators of corruption perception rather than using a measure derived from tariff data. The reason is that the latter might depend too strongly on the particulars of the tariff-setting model as well as the quality of the data used. A schematic overview of the methodology and results of the empirical literature can be found in appendix 2.A.

2.3.3 Rent-shielding

In contrast to the rent-seizing idea, Maggi and Rodríguez-Clare (1998) see RIAs purely as a way for the government to limit the power of lobbies and eliminate certain sources of rents. Firstly, the enhanced competition of the rent-destruction effect eliminates sources of rents. Additionally, by encapsulating policy in international agreements the direct control of the government is lowered, which also eliminates potential rent-seeking behavior. In both cases, the effect of an agreement is to lower corruption.

With respect to the reciprocal effect of corruption on the willingness to join an agreement, this model also predicts an inverted U-shape. If the level of corruption is too high, too many people would lose when an agreement is signed for a majority to support it. Vice versa, if the government cared only about welfare, lobbies would not be able to exert any pressure to start with. Only at intermediate levels would an increase in corruption provide an additional incentive to government officials to combat corruption using the trade agreement (Maggi and Rodríguez-Clare, 1998).

2.4 Data

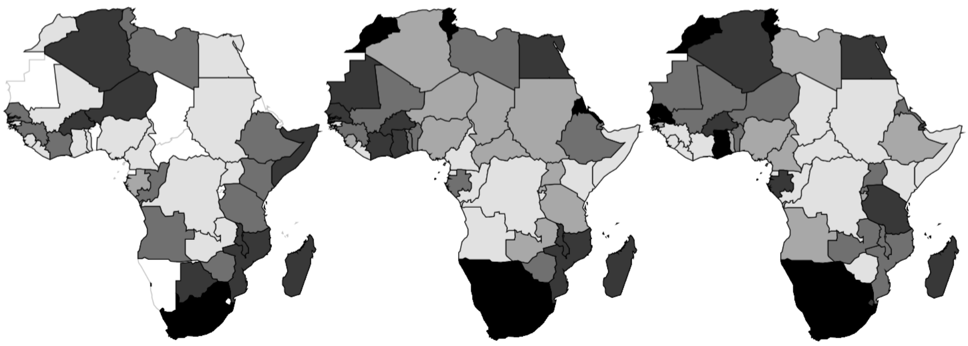
To test the link between corruption and integration, data was collected on the founding of, and accession to regional integration agreements in Africa. This was done using the Regional Integration Knowledge System (UNU-CRIS, 2006), and the webpages of the regional trade agreements themselves. Twelve free trade agreements and customs unions were incorporated, which taken together cover 53 African countries:

Western and Northern Africa:	
AMU	African Maghreb Union;
CEN-SAD	Community of Sahel-Saharan States;
ECOWAS	Economic Community of West African States;
GAFTA	Greater Arab Free Trade Area;
UEMOA	West African Economic and Monetary Union.
Central Africa:	
CEPGL	Economic Community of the Great Lakes;
UDEAC	Central African Economic and Monetary Union;
ECCAS	Economic Community of Central African States.
Eastern and Southern Africa:	
EAC	Eastern African Cooperation;
COMESA	Common Market for Eastern and Southern Africa;
SACU	Southern African Customs Union;
SADC	Southern African Development Community.

This dataset was used to create two dependent variables to be used in the unilateral and bilateral regressions. $\text{RIAuni}_{x,t}$ indicates whether country x signed an agreement in year t or any of the previous four years. This indicator was constructed in five year intervals from 1985 to 2010, for a total of 318 observations. On average, 34% of the countries signed an agreement within a five year interval, more or less equally divided over the 25 years of the sample.

$RIAbi_a^b$ signals whether countries a and b are members of any of the listed agreements in the year 2010. Because the index is symmetrical, each country-couple is covered once giving us $53 * 52/2 = 1378$ observations. Because of the plethora of African trade agreements, half of the countries-couples covered are members of the same agreement.

2.4.1 Measuring corruption



(a) ICRG - 1985 **(b)** WGI - 1998 **(c)** CPI - 2007

Figure 2.1: Map of the used corruption values, ranging from little (dark) to a lot of corruption (light).

The two best known subjective indicators of corruption are the Corruption Perceptions Index (CPI) created by Transparency International (2008) and the Worldwide Governance Indicators' control of corruption index (WGI) constructed by Kaufmann et al. (2010) The problem with both composite indices is that they only go back to the mid-nineties (1996 for the WGI index and 1995 for the CPI index). In addition, their earliest estimates are based on a limited number of sources and are therefore much more uncertain than their more recent ones. How much this matters can be seen from the WGI index that reports an estimate of its measurement error: the average prediction error for the African continent drops to half its value from

1996 to 2008 (Kaufmann et al., 2010).

While the WGI index covers every African country from 1998 onwards, the initial estimates of CPI suffer from a strong selection bias. As Treisman (2007) points out, the first corruption surveys were aimed at countries that were important to the international markets. Small and/or highly corrupt countries were less likely to be surveyed. To avoid this problem, we used the values of the first year where all 53 countries were covered: 1998 (WGI) and 2007 (CPI). Lastly, it should be remarked that the way the CPI index is constructed does not allow for comparisons over time (Transparency International, 2012). However, this should not pose a problem in cross-sectional analyses.

The third indicator of corruption is that of the International Country Risk Guide (ICRG). Its main advantage over CPI and WGI is that it goes back to 1984, allowing the use of pre-dated values to avoid possible endogeneity problems. However, the 1985 values only cover 34 countries of the 53 countries in our sample.

For all three indices, high values correspond to low levels of corruption. In other words, the rent-seizure, rent-destruction and rent-shielding hypotheses predict a negative sign for the coefficient on the corruption variables. To facilitate comparability, all indices are rescaled such that their worldwide values range from zero (high corruption) to one (no corruption).[4] The smallest value for an African country is zero for all the indicators, while the maximum equals one (ICRG), 0.64 (WGI) and 0.5 (CPI).

Apart from using predated values (when possible) we also control for endogeneity using several instruments for corruption and institutional development. A number of instruments were considered: tropics and primary export intensity (Sachs and Warner, 1995), slave exports (Nunn, 2009), settler mortality (Acemoglu et al., 2001), former British colonies (Landes, 1998) and ethnolinguistic fractionalization (Mauro, 1995).

The only supply-side instrument is the primary export intensity. A high fraction of primary exports is likely to go hand in hand with high rents which would in

[4] $Corr = (Corr^{\star} - \min_{\text{world}}(Corr^{\star}))/(\max_{\text{world}}(Corr^{\star}) - \min_{\text{world}}(Corr^{\star}))$.

turn raise the probability of rent-seeking behavior and corruption. The remaining variables all affect the demand side of corruption, mostly through weakened institutions. Tropical diseases and a high settler mortality rate would encourage colonists to set up short term, maximum rent-extracting institutions rather than invest in good governance. High ethnolinguistic fractionalization and a history of slave exports are likely to undermine public trust in the government even in the long term. Finally, it is argued in the literature that British colonizers left behind superior institutions relative to their French, Portuguese or Spanish counterparts. For an in-depth discussion of how these instruments affect the level of corruption we refer to the papers listed.

Because the first stage regressions only have 53 observations and there are many missing data points among the instruments variables, only two best scoring instruments were used for each corruption indicator. Appendix 2.B lists the results of the first stage regression of the corruption indicators on the two best-scoring instruments. The F-statistic of the regression on CPI and WGI is at, or above the Staiger-Stock rule of thumb of 10. On the other hand, the instruments do a much worse job capturing the variation in ICRG.

Showing the exogeneity of the instruments is fairly straightforwards. Tropics is a geographical feature, while settler mortality, former British colonies and slave exports were all determined (over) a century ago. The most recent variables are the primary export intensities in the 70s and ethnolinguistic fractionalization in the 60s, but both still predate the trade agreements by as much as 20 years.

It is unclear how the variables would influence the willingness to form trade agreements other than through their effect on corruption or possibly the control variables. A high values of tropics might indicate that countries lie close together, but this is already controlled for by the distance variable. British colonies might be more likely to form trade agreements, but this is captured by the shared colonial history variable. Countries that are ethnolinguistically fractionalized are countries whose actual borders don't match their 'natural borders'. It could be argued that they might seek to address this by closing trade agreements. Nevertheless, this variable was not selected as instrument and neither was the primary export intensity. Finally, it is unclear how the settler mortality rate and slave exports would directly affect the probability of a trade agreement.

2.4.2 Control variables

Several control variables are taken into consideration, most of them coming from the aforementioned papers studying the determinants of regional integration. They can be divided into three groups: geographical, economic or political. Their main characteristics, including a correlation table, can be found in appendix 2.C. The geographical variables are:

- *Landlocked* is expected to have a positive sign, since the countries that are cut off from the world markets would be more willing to join RIAs.

- The more *remote* countries are from the rest of the continent, the lower the opportunity cost of them signing a RIA. The variable is computed as described in Baier and Bergstrand (2004a).

- The remaining two geographical variables are indicators of the geographical distance between countries. For each variable, we expect that the closer the countries are, the stronger their inclination to form a RIA is. *Natural* is the inverse of the distance between capitals and *adjacent* indicates whether the two countries neighbor one another.

The economic variables come from the Penn World Tables 8.0 (Feenstra et al., 2013):

- The first economic variable is the difference in the capital-labor ratios between countries (*DKL*). The bigger the difference, the higher the expected trade creation effects, and the more likely an agreement is. We used total population as a proxy for labor.

- Similar to the remoteness variable, the difference in capital-labour ratios with the rest of the continent (*DROWKL*) was also included (Baier and Bergstrand, 2004a).

- Following Baier and Bergstrand (2004a), the level of *GDP* is included to control for economies of scale. Alternatively, Wu (2006) uses GDP per capita to control for the level of economic development. However, using per capita GDP instead of GDP did not alter the results in any way (available upon request).

- The last economic variable measures the difference in GDP's of both countries. The smaller the difference in GDP, the larger the net welfare effect of the trade agreement should be (Baier and Bergstrand, 2004a).

Finally, in addition to the level of corruption there are two other political variables that are taken into consideration:

- *Colony* is a dummy variable that is one when the countries have an identical colonial background. Colonial history is also highly correlated with the official language of a country and could be interpreted as a proxy for it.

- *Polity* indicates the level of democracy versus autocracy in a country: -10 being a completely totalitarian regime and 10 a completely democratic one. Mansfield et al. (2002) find that it has a positive effect on the likelihood to enter a RIA. The data comes from the Polity IV dataset (Marshall et al., 2014).

2.5 Econometric specification

When estimating the effect of corruption on RIAs two issues require special attention. Firstly, the agreements are not bilateral but involve more than two partner countries. Secondly, there is the problem of endogeneity: the theoretical models all posit either a positive or a negative effect of RIAs on corruption. In case of ICRG, we can use predated values as a first step to control for this. For WGI and CPI we can only use instrumental variables.

The literature proposes two different econometric models to determine the factors that drive regional integration attempts. We label these the unilateral and bilateral approaches.

Wu (2006) uses the *unilateral approach* (equation 2.1), regressing whether or not country x entered a RIA in year t on the characteristics of that country in that year. For example, this can be used to test whether landlocked nations are more likely to join a RIA or how the level of GDP affects the willingness to join.

$$Pr(RIAuni_{x,t} = 1) = \Phi[\,Corr_{x,t}, Corr^2_{x,t}, GDP_{x,t} \times Corr_{x,t}, \bar{Z}_{x,t}\,] \tag{2.1}$$

The probability that country x is a member of a trade agreement at time t is a function of a linear combination of its level of corruption, its squared values as well as an interaction with GDP. $\bar{Z}_{x,t}$ contains additional country characteristics that are controlled for, such as its level of democracy. When estimating a probit model the function Φ is the standard normal cumulative distribution function.

The main drawback of the unilateral approach is that the characteristics of the partner countries cannot be taken into account. As a result, this model performs remarkably bad when used to explain African integration: with the exception of *landlocked* at the 10% level, none of the explanatory variables are significant (appendix 2.D).

The *bilateral approach* on the other hand does allow the characteristics of one partner country to be taken into account. It entails regressing whether or not two countries have formed a RIA on the characteristics of both countries. There are two ways of estimating the effect of corruption in the bilateral regressions: using average levels or individual levels.

The *average-levels bilateral approach* (equation 2.2) is used in Baier and Bergstrand (2004a), Endoh (2006), Márquez-Ramos et al. (2011) and Arcand et al. (2011). The probability that country a and b are members of the same agreement is a function of the linear combination of their average level of corruption, its squared value and other characteristics of the country-couple (Z_a^b):

$$Pr(RIAbi_a^b = 1) = \Phi\left[\frac{Corr_a + Corr_b}{2}, \left(\frac{Corr_a + Corr_b}{2}\right)^2, Z_a^b\right] \tag{2.2}$$

The underlying assumption of this approach is that an increase in corruption has the same effect regardless of the characteristics of the partner country.

The second way in which corruption can be added is by including the corruption index of both countries separately, similar to how Mansfield et al. (2002) treat the level of democracy. This *individual-levels bilateral approach* (equation 2.3) allows us to see whether the effect of corruption changes depending on the relative size of the country, as suggested by the model of Grossman and Helpman (1995).

$$Pr(RIAbi_a^b = 1) = \Phi[Corr_a, Corr_a^2, Corr_b, Corr_b^2, Z_a^b] \tag{2.3}$$

It should be clear from equations 2.2 and 2.3 that countries *a* and *b* enter the equations symmetrically and the labels can be switched around with no apparent effect. For this reason, each country couple is only covered once, regardless of the order in which the countries are listed. In other words, if *n* countries are covered, the dataset will contain a maximum of $C_2^n = \frac{n(n-1)}{2}$ observations. In contrast, in for example the Grossman and Helpman (1995) model only one country's level of corruption will influence the probability of an agreement because a positive welfare effect would ensure that the partner country is always willing to join. In their model, the deciding factor is the relative size of both countries, which is why the countries were labeled in such a way that country *a* is the country with the highest GDP of the country-couple. In other words, the variation in the effect of corruption between country *a* and *b* can be attributed to differences in relative size. To be clear, this did not alter the selection of countries, but merely made the interpretation of the variables more meaningful.

A problem with using the bilateral approach to analyze African trade agreements is that most agreements involve more than two partner countries and that many are overlapping. The former means that we have to take into account that the decision of two countries to enter into an agreement crucially depends on the decision of the other partner countries in that agreement, while the latter ensures that this pattern of interaction effects becomes very complex. As is well known, ignoring any form of heteroscedasticity in a probit model could render the estimations inconsistent (Greene, 2002). We control for this by correcting the standard errors for multiple, non-nested clusters using the methodology outlined in Cameron et al. (2006). While this methodology is relatively simple to implement, a rise in the number of cluster variables exponentially increases the number of additional regressions that have to be run.[5] For this reason, the twelve trade agreements were grouped into six categorical variables such that no overlapping agreements are captured by the same variable. This significantly lowered the computational burden, without affecting the results. Take for example the categorical variable R_2: it is one if both countries are a member of SADC, two if they are members of GAFTA, and zero otherwise.

R_1: SACU-ECOWAS-EAC-AMU	R_4: CENSAD-CEPGL
R_2: SADC-GAFTA	R_5: UEMOA-CEMAC
R_3: ECCAS	R_6: COMESA.

In order to control for both the heteroskedasticity as well as the endogeneity, we used a control function approach. The endogenous corruption variables were first regressed on their instruments and the remaining exogenous dependent variables. By including the error term of this first-stage regression (ε) in the probit model outlined above, the endogenous part of corruption can be controlled for (Woolridge, 2005). As mentioned in section 2.4.1 when discussing the instrumental variables, the number of instruments was limited to two for each corruption variable, preventing the first-stage regressions from losing too many observations. While the instruments have a relatively low R^2, they are jointly significant for all three corruption indicators.

[5]To correct the standard errors of one regression with n cluster variables, an additional $2^n - 1$ estimations have to be run. If we use the 12 dummy variables that are one when both countries are member of a certain RIA as the cluster variables, then each regression would require 4,095 extra estimations.

The standard errors of the control function approach were subsequently computed using bootstrapping.

Bootstrap iteration:

1. Draw a new bootstrap sample (with replacement);
2. Run first-stage regression and collect error term;
3. Run the probit model, including the first-stage error term as dependent variable;
4. Correct standard errors for non-nested clusters;
5. Save standard errors and repeat from step 1.

To check the robustness of our results, we also used the two-step instrumental variable probit estimator (Newey, 1987). However, the standard errors using the ivprobit approach could not be corrected for the heteroscedasticity caused by the overlapping trade agreements.

2.6 Empirical results

2.6.1 Average-levels bilateral regressions

Table 2.2 replicates the Baier and Bergstrand (2002) and Mansfield et al. (2002) regressions with the addition of the average level of corruption and its squared value (equation 2.2). In columns one and two, corruption is measured using ICRG's index, three and four use the Worldwide Governance Indicators and five and six use the Corruption Perception Index.

In contrast to table 2.1, the regressions in table 2.2 are corrected for the multilateral aspect of trade agreements. As a result, the capital-labor ratio and its deviation with the rest of the world becomes significant. In line with the prediction and findings of Baier and Bergstrand (2004a), the higher the difference in the capital-labor ratio between two countries, the greater the advantage of lowering barriers to trade and

Table 2.2: Average-levels probit regression

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.530***	1.422***	1.437***	1.334***	1.541***	1.448***
	(12.78)	(9.95)	(13.58)	(11.01)	(12.30)	(10.48)
Remote	4.453***	4.481***	4.839***	4.707***	5.926***	5.692***
	(4.17)	(3.63)	(3.72)	(3.10)	(3.99)	(3.55)
Adjacent	–	0.870***	–	0.865***	–	0.867***
		(15.28)		(13.96)		(9.83)
Landlocked	–	0.0689	–	0.0525	–	0.0648
		(0.43)		(0.27)		(0.41)
GDP_{av}	−0.0473	−0.0671	−0.0672	−0.105	−0.093	−0.133*
	(−0.68)	(−0.78)	(−0.79)	(−1.13)	(−1.19)	(−1.65)
GDP_{diff}	0.0838**	0.114***	0.0787**	0.102***	0.103***	0.124***
	(2.12)	(2.90)	(1.99)	(2.90)	(3.07)	(4.59)
DKL	0.0276***	0.0293**	0.0284	0.0420*	0.0638***	0.0817***
	(2.98)	(2.57)	(1.61)	(1.68)	(4.74)	(3.91)
DROWKL	−0.057	−0.0653**	−0.065	−0.0689***	−0.168***	−0.175***
	(−1.16)	(−2.20)	(−1.44)	(−2.76)	(−3.03)	(−4.82)
Colony	–	0.340***	–	0.392***	–	0.338***
		(7.59)		(6.25)		(4.41)
$Polity_{av}$	–	−0.0307***	–	0.00339	–	0.0208**
		(−3.56)		(0.24)		(2.19)
$Corruption_{av}$	−0.244	−0.355	−0.388	−0.251	−7.090***	−6.258***
	(−0.30)	(−0.53)	(−0.13)	(−0.09)	(−4.24)	(−3.94)
$Corruption^2_{av}$	−1.545*	−1.446**	−5.180	−5.768	−0.874	−3.630
	(−1.81)	(−1.99)	(−0.97)	(−1.24)	(−0.15)	(−0.63)
LR[(a)] corr	$5.05e-08$	$2.24e-7$	$2.33e-9$	$4.37e-9$	$1.72e-22$	$1.09e-21$
Constant	−23.98***	−25.48***	−27.45***	−27.22**	−34.81***	−33.59***
	(−2.93)	(−2.67)	(−2.62)	(−2.22)	(−2.95)	(−2.62)
Observations	992	992	1128	1127	1128	1127

Probit regression of RIAbi on corruption and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. [(a)]p-value of the likelihood ratio tests on the joint significance of the corruption variables. *, **, *** indicates significance at the 10%, 5% and 1% level.

the more likely they enter into an agreement. Similarly, as the difference with the rest of the continent decreases, the opportunity cost of the agreement decreases and the probability of joining increases. Nevertheless, the effect of both factors is small when compared to distance (natural and remoteness), or having a similar colonial history.

When using the ICRG index, a number of other variables (e.g. GDP_{av}, $Polity_{av}$) also change signs or significance when compared to the regressions using CPI or WGI. This is caused by selection effects (cf. infra): as is evident from the change in the number of observations, ICRG is not available for all countries. In an attempt

Table 2.3: Average-levels probit regression with control function

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.518***	1.420***	1.422***	1.324***	1.482***	1.397***
	(12.09)	(9.24)	(12.34)	(11.24)	(13.30)	(11.30)
Remote	5.482***	6.096***	5.860***	6.351***	7.077***	6.467***
	(5.24)	(4.60)	(5.20)	(3.52)	(3.84)	(3.16)
Adjacent	–	0.915***	–	0.828***	–	0.840***
		(11.11)		(13.65)		(8.83)
Landlocked	–	0.350**	–	0.183	–	0.125
		(2.32)		(0.89)		(0.65)
GDP_{av}	−0.0133	0.0123	−0.0304	−0.0443	−0.0718	−0.116
	(−0.23)	(0.15)	(−0.40)	(−0.43)	(−0.79)	(−1.18)
GDP_{diff}	0.0695*	0.0935**	0.0608*	0.0721*	0.106***	0.126***
	(1.90)	(2.30)	(1.76)	(1.89)	(2.93)	(3.97)
DKL	0.0131	0.0155	0.0374**	0.0589**	0.0917***	0.108***
	(1.35)	(1.06)	(2.30)	(2.27)	(3.86)	(3.20)
DROWKL	−0.139*	−0.211***	−0.0751	−0.0951***	−0.233***	−0.226***
	(−1.88)	(−4.04)	(−1.57)	(−4.00)	(−6.18)	(−6.42)
Colony	–	0.353***	–	0.438***	–	0.337***
		(7.93)		(6.13)		(4.27)
$Polity_{av}$	–	−0.0242***	–	0.0196	–	0.0349***
		(−2.98)		(1.52)		(2.74)
$Corruption_{av}$	−3.699***	−5.351***	−3.241	−5.526**	−10.52***	−9.413***
	(−7.70)	(−10.51)	(−1.29)	(−2.35)	(−15.00)	(−38.32)
$Corruption^2_{av}$	−1.114	−0.901	−4.209	−3.297	−2.554	−3.686
	(−1.55)	(−1.45)	(−0.84)	(−0.80)	(−0.37)	(−0.60)
LR[a] corr	$1.33e-19$	$9.77e-69$	$4.34e-12$	$2.81e-7$	$6.44e-88$	0
Error term ε	3.250***	4.698***	2.873***	4.822***	4.679**	3.795
	(2.78)	(4.43)	(2.68)	(4.76)	(2.36)	(1.60)
Constant	−31.15***	−37.14***	−35.33***	−39.67***	−44.23***	−39.89**
	(−3.77)	(−3.57)	(−3.78)	(−2.73)	(−3.02)	(−2.44)
Observations	970	970	1125	1124	1097	1097
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

Probit regression of RIAbi on corruption, the error term from the first stage regressions (ε) and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. [a]p-value of the likelihood ratio tests on the joint significance of the corruption variables. *, **, *** indicates significance at the 10%, 5% and 1% level.

to limit selection problems, the average was also computed when only one corruption indicator was available: $Corr_{av} = Corr_a$ if $Corr_b$ is missing and vice versa. Without this correction, the number of observations is reduced by more than half and a number of variables (for example having a similar colonial history) lose their significance (table 2.4).

Depending on the corruption indicator used, the significance of corruption can change drastically, however this is entirely due to the multicollinearity problem between corruption and its square: the correlation exceeds 90% for each indicator. Jointly, the corruption indicators are always significant at more than 1% levels, as shown by the likelihood ratio tests.

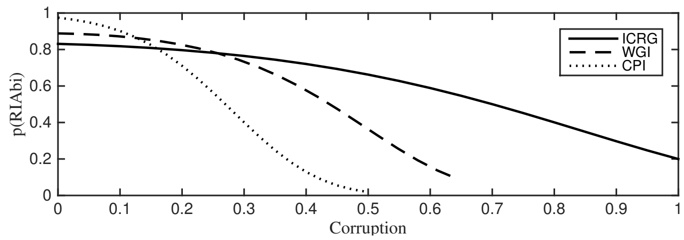
Using a control function to check for endogeneity does not change the overall conclusion of the results. While a few of the individual corruption variables change significance, the overall significance of corruption remains unaffected. The biggest change happens to the parameter on *adjacency* which loses its significance in case of CPI and WGI. Similarly, a number of the economic variables change significance when ICRG is used. Nevertheless, the parameter on the control function (ε) tends to be significant, which would indicate the need to control for endogeneity.

Figure 2.2 summarizes the effect of corruption on the willingness to enter into an agreement by plotting out the marginal effect of an increase in corruption. It clearly shows that as corruption gets worse (or the corruption indexes decrease), the probability of joining is raised, regardless of whether or not we control for endogeneity. Using ivprobit instead of a control function also does not affect the results (appendix 2.E). While the individual parameters of some of the corruption variables change considerably, the overall marginal effect of corruption is very similar to both the simple probit model and the control function approach. Only the instrumented ICRG regression finds evidence of a small reversal at high levels of corruption.

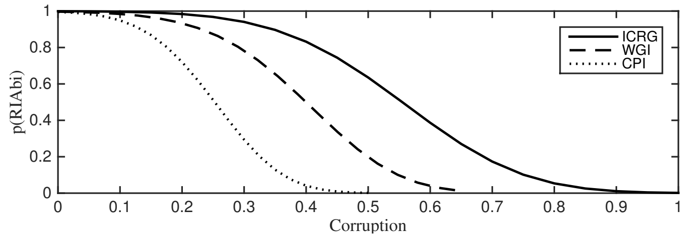
2.6.2 Individual-levels bilateral regressions

Including the corruption indicators separately allows us to check whether there are asymmetries in the effect of corruption depending on the relative size of the countries (cf. Grossman and Helpman, 1995). Overall, the coefficients on the control variables change little. The biggest changes occur when using the ICRG index, but as explained this can be attributed to the notable drop in observations.

Corruption and corruption squared are jointly significant for both countries at more than 1% in the probit (table 2.4), control function (table 2.5) and ivprobit model (appendix 2.E). Figures 2.3 and appendix 2.E plot the marginal effects of a change



(a) Probit regressions (table 2.2)



(b) Probit with control function (table 2.3)

Figure 2.2: Marginal effect of corruption and GDP: average-levels
The marginal effect of average corruption on the probability of joining a RIA, keeping all other variables at their mean values. The maximum values of WGI and CPI in Africa are 0.62 and 0.5, while ICRG's values range between 0 and 1. Higher index values correspond to lower levels of corruption.

in corruption on the probability of joining for both countries for the different regression models and indicators of corruption. They show that while the parameters on corruption and its interaction effects are again relatively unstable, their combined effect is much less divergent.

The effect of an increase in the largest country in terms of GDP (country A) is similar over the three corruption indicators and estimation methods. Overall the probability of joining rises as corruption becomes worse. However, when a control function is used, the effect of an increase in the ICRG corruption variable is very small, but still significant. At the same time, the coefficients on the endogenous part of corruption (ε_a and ε_b) are never significant, indicating that the (already lagged) ICRG variable should not be further corrected for endogeneity. Moreover, when ivprobit is used, the effect of an increase in ICRG becomes larger again. The instability of the results using ICRG can be explained in part by the substantial reduction in the number of observations in these regressions.

The parameters on corruption of the smallest country (country B) differ more strongly

Table 2.4: Individual-levels probit regression

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	2.026***	1.913***	1.436***	1.297***	1.547***	1.409***
	(10.48)	(10.24)	(12.59)	(9.20)	(12.42)	(9.34)
Remote	8.247***	8.179***	5.084***	5.865***	5.995***	6.161***
	(10.19)	(9.00)	(3.80)	(3.32)	(4.05)	(3.59)
Adjacent	–	0.868***	–	0.697***	–	0.705***
		(11.14)		(10.47)		(32.41)
Landlocked	–	0.0293	–	0.189	–	0.148
		(0.14)		(0.94)		(1.00)
GDP_a	0.171***	0.247***	0.0392**	0.0562**	0.0574***	0.0684**
	(3.89)	(7.37)	(1.98)	(2.01)	(2.84)	(2.37)
GDP_b	0.142	0.0893	−0.00444	0.0683	−0.0389	0.00695
	(1.49)	(0.64)	(−0.09)	(1.39)	(−0.87)	(0.17)
DKL	0.0233**	0.0190	0.0391	0.0617*	0.0721***	0.109***
	(2.45)	(1.06)	(1.62)	(1.81)	(4.21)	(4.28)
DROWKL	0.111*	0.0882*	−0.0781	−0.0652**	−0.160***	−0.162***
	(1.85)	(1.89)	(−1.64)	(−2.22)	(−2.75)	(−3.65)
Colony	–	0.102	–	0.409***	–	0.358***
		(1.45)		(5.58)		(3.59)
$Polity_a$	–	−0.0268	–	−0.0147***	–	−0.0137***
		(−1.52)		(−4.52)		(−2.81)
$Polity_b$	–	−0.0160	–	0.0255	–	0.0302**
		(−0.88)		(1.49)		(2.29)
$Corruption_a$	−1.122	−1.030	−0.403	−1.104	−4.503***	−5.811***
	(−0.69)	(−0.62)	(−0.21)	(−0.53)	(−5.29)	(−7.00)
$Corruption_a^2$	−0.573	−0.578	−2.727	−2.204	0.516	3.189
	(−0.33)	(−0.31)	(−1.00)	(−0.77)	(0.21)	(1.24)
$Corruption_b$	0.104	−0.0685	1.463	0.237	−0.626	−1.298
	(0.10)	(−0.07)	(1.51)	(0.17)	(−0.34)	(−0.60)
$Corruption_b^2$	−0.699	−0.446	−4.668**	−4.563*	−5.211	−5.025
	(−0.44)	(−0.32)	(−2.23)	(−1.83)	(−1.52)	(−1.22)
$Wald^{(a)}$ $corr_a$	$1.28e-8$	$1.55e-5$	$6.30e-6$	$3.96e-4$	$4.36e-16$	$6.88e-17$
$Wald^{(a)}$ $corr_b$	$1.22e-4$	0.00181	0.00405	$1.19e-12$	$5.86e-4$	$1.15e-7$
Constant	−53.37***	−54.34***	−29.86***	−37.83***	−35.67***	−38.56***
	(−8.46)	(−7.72)	(−2.82)	(−2.70)	(−3.01)	(−2.79)
Observations	465	465	1128	1035	1128	1035

Probit regression of RIAbi on corruption and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. $^{(a)}$p-value of the wald tests on the joint significance of the corruption variables of each country. *, **, *** indicates significance at the 10%, 5%, and 1% level.

depending on the estimation model. When a simple probit model is used, the pattern does not differ from that of the larger country: an increase in corruption raises the probability of a RIA. Using a control function even reverses the effect in the case of ICRG, but this is not the case in the ivprobit model. When an ivprobit model is used, we see signs of a reversal only at very high levels of corruption for ICRG and WGI,

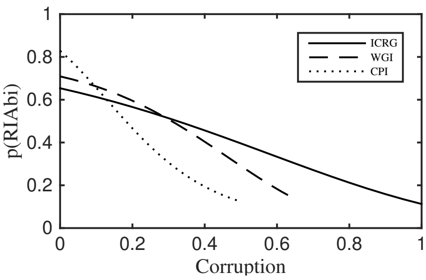
Table 2.5: Individual-levels probit regression with control function

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.646***	2.087***	1.574***	1.415***	1.956***	2.320***
	(7.75)	(10.00)	(11.09)	(9.68)	(9.60)	(11.35)
Remote	7.804***	8.487***	5.867***	6.661***	7.968***	9.546***
	(5.12)	(4.76)	(4.30)	(3.37)	(3.60)	(4.12)
Adjacent	–	0.471***	–	0.397***	–	−0.427***
		(3.52)		(3.36)		(−3.05)
Landlocked	–	−0.512**	–	0.234	–	0.00132
		(−2.53)		(1.03)		(0.01)
GDP_a	0.432***	0.471***	0.144***	−0.00553	−0.0357	0.0237
	(6.97)	(6.69)	(6.03)	(−0.19)	(−0.98)	(0.65)
GDP_b	−0.313	−0.213	0.0739	0.0827	−0.0293	0.0451
	(−1.43)	(−1.28)	(1.52)	(1.51)	(−0.60)	(0.79)
DKL	−0.164***	−0.0247	0.0168	0.0809*	−0.00637	0.110***
	(−4.71)	(−0.63)	(0.42)	(1.90)	(−0.21)	(3.36)
DROWKL	0.0902	0.157	0.0963*	0.0565	0.0389	−0.0254
	(0.46)	(0.74)	(1.67)	(1.51)	(0.54)	(−0.48)
Colony	–	−0.159*	–	0.510***	–	0.0401
		(−1.79)		(5.22)		(0.32)
$Polity_a$	–	0.000569	–	−0.0134*	–	−0.00553
		(0.03)		(−1.79)		(−0.57)
$Polity_b$	–	0.0194	–	0.0268*	–	0.0195
		(0.40)		(1.66)		(0.95)
$Corruption_a$	−1.611	−0.331	−1.436	−4.075	−2.100	−6.960**
	(−1.60)	(−0.29)	(−0.57)	(−1.24)	(−0.80)	(−2.48)
$Corruption_a^2$	−0.770	−1.884	−2.506	0.569	−0.884	4.694
	(−0.47)	(−1.10)	(−0.84)	(0.16)	(−0.22)	(1.03)
$Corruption_b$	1.654*	−5.244***	−2.011**	−1.976	−12.73***	−9.800***
	(1.66)	(−4.78)	(−2.57)	(−1.00)	(−7.50)	(−4.67)
$Corruption_b^2$	−1.683	6.236***	1.831	−2.112	7.924**	10.46***
	(−0.84)	(3.23)	(0.94)	(−0.74)	(2.04)	(3.18)
$Wald^{(a)}$ $corr_a$	$1.27e-4$	0.0439	$1.65e-6$	$2.71e-4$	0.0327	$4.78e-5$
$Wald^{(a)}$ $corr_b$	0.00374	0.00211	0.0047	$4.07e-12$	$6.76e-10$	$4.81e-10$
Error term ε_a	0.488	0.143	1.776***	2.286**	−1.955	0.304
	(0.50)	(0.14)	(2.66)	(2.17)	(−1.59)	(0.21)
Error term ε_b	0.194	0.889	−1.312	0.933	5.294***	1.079
	(0.09)	(0.42)	(−1.53)	(1.28)	(4.56)	(0.89)
Constant	−51.36***	−53.35***	−36.16***	−42.14***	−46.37***	−57.46***
	(−4.54)	(−3.97)	(−3.43)	(−2.76)	(−2.72)	(−3.21)
Observations	339	356	1023	951	710	687
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

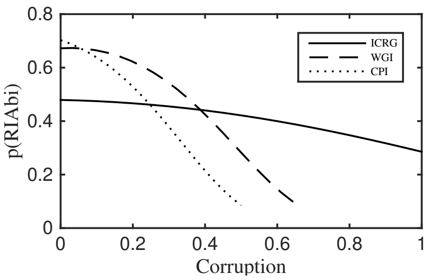
Probit regression of RIAbi on corruption, the error terms from the first stage regressions (ε_a and ε_b) and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. $^{(a)}$p-value of the Wald tests on the joint significance of the corruption variables. *, **, *** indicates significance at the 10%, 5% and 1% level.

consistent with the findings of for example Arcand et al. (2011).[6] The fact that the

[6] Including an interaction term between corruption and GDP as in Arcand et al. (2011) also does not change this conclusion, but the interaction term adds little explanatory power. The results have been omitted to save space, but are available on request.

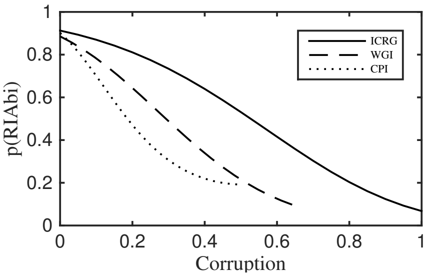


(a) Country A

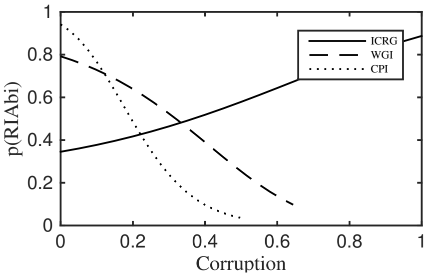


(b) Country B

Probit regressions (table 2.4)



(c) Country A



(d) Country B

Probit with control function (table 2.5)

Figure 2.3: Marginal effect of corruption and GDP: individual-levels
The marginal effect of average corruption on the probability of joining a RIA, keeping all other variables at their mean values. The maximum values of WGI and CPI in Africa are 0.62 and 0.5, while ICRG's values range between 0 and 1. Higher index values correspond to lower levels of corruption.

larger country's level of corruption is consistently negative and significant falls in line with the predictions of Grossman and Helpman (1995). In their model only the larger country can have a negative effect on its welfare, making the preference for rents of its government a deciding factor.

2.6.3 Robustness

We checked to what extent the differences in the effect of corruption when using different corruption indicators were caused by differences in the sample size. To that end, the same sample size was kept constant for all regressions (available upon request). The effect of corruption on the probability of an agreement remained unaffected by the change in sample size. The control variables are now almost completely unaffected by the choice of corruption indicator: the only parameters that still vary are those on variables that are economically insignificant (e.g. polity). Comparing the instrumented with the uninstrumented result when the number of observations is kept constant shows that the effect of controlling for endogeneity is relatively limited. Overall, the size of the coefficient on the average corruption level rises in case of ICRG and WGI, but not when using CPI. In the separate level regressions, the effect of the corruption in the largest country also increases (but that of the smaller country tends to become smaller). An increase in the size of the parameters when controlling for endogeneity suggests that corruption has decreased after the agreement was signed, which would match the rent-shielding and rent-destruction hypotheses. Nevertheless, the effect is very small and changes depending on the size of the country and corruption indicator used.

In the analyses presented above only the two strongest instruments were used in an attempt to limit the number of missing variables. However, selecting instruments based on which one gives the best fit could potentially lead to pre-test biases in the results. To make sure this did not affect the results, the analysis was repeated using the six instruments listed in the data section (2.4.1). While this severely reduced the number of observations, the effect of corruption remained unaffected (available upon request).

The monotonic effect of corruption

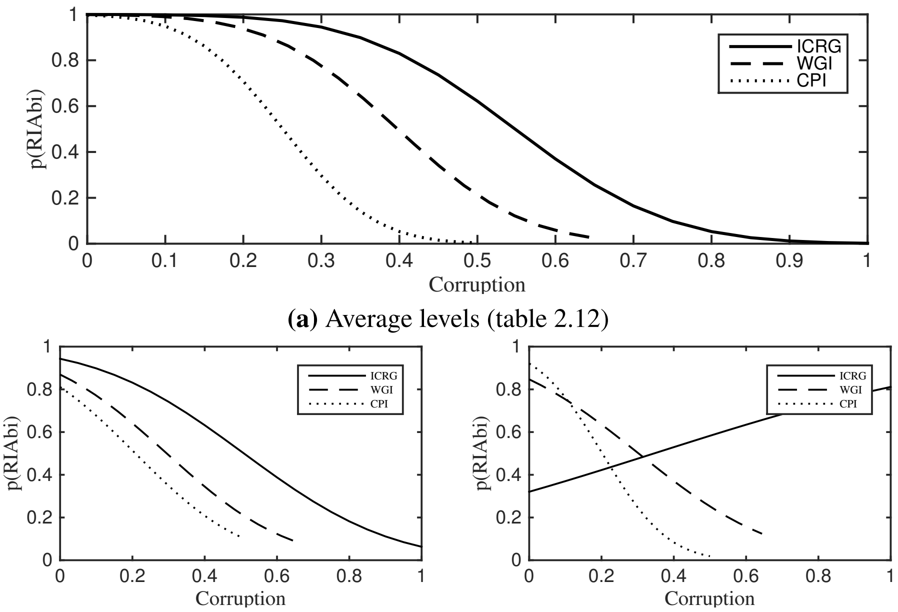
The rent-destruction and rent-shielding hypotheses predicted that the effect of corruption on the probability of an agreement would follow an inverted U-shape pattern. However, in more than five out of six cases this pattern failed to emerge. Moreover, the high correlation between corruption and its square value lead to multicollinearity problems that could distort the results. As a robustness check, the model was re-estimated using only the level of corruption.[7]

Appendix 2.F list the parameter values estimation using a control function, while figure 2.4 plots the marginal effects. Apart from a minimal change in the curvature

[7] The results did not change when only the squared values of corruption were used. Similarly, the probit and IV probit results follow the same pattern, all are available upon request.

of some of the figures, the results are almost identical. In the average level estimations corruption is negative and significant at the 1% level regardless of which indicator of corruption is used. When including the country characteristics separately, ICRG is significant at the 10% level only for the largest of the two countries, but the results also indicate that the correction for endogeneity is not necessary. Without the controls for endogeneity it is significant at the 1% level. WGI and CPI remain highly significant for both countries and the correction for endogeneity is shown to be necessary.

The fact that the non-monotonic relationship between corruption and trade agreements fails to occur could be seen as evidence that the rent-destruction and rent-shielding effects do not come into play. However, the regressions performed here at not the best way of discerning between the different models and there are many other possible reasons (e.g. measurement errors in the corruption variables) why this pattern did not emerge.



(a) Average levels (table 2.12)

(b) Separate: Country A (table 2.13)

(c) Separate: Country B (table 2.13)

Figure 2.4: Marginal effects when regressing without the squared values of corruption
The marginal effect of average corruption on the probability of joining a RIA, keeping all other variables at their mean values. The maximum values of WGI and CPI in Africa are 0.62 and 0.5, while ICRG's values range between 0 and 1. Higher index values correspond to lower levels of corruption.

2.6.4 Relevance of corruption in explaining RIAs

While the regressions above show that corruption has a significant impact on the decision to join a trade agreement in Africa, they have not established the relevance of this effect. To this end, table 2.6 compares the predictive power of various models. Because missing data in some variables changes the estimation sample, this comparison has to be made made model-by-model. Especially when ICRG is added, the number of observations drops significantly.

After listing the models and the table and column where they can be found, table 2.6 indicates how many of the observations were correctly classified by each model. An observation is correctly classified if RIAbi is one (zero) and the predicted probability is higher (lower) than 0.5. To capture the overall change at all probabilities instead of only at the 0.5 point, the table also lists the average prediction error. The latter is defined as the Euclidian distance between RIAbi and the predicted probability (equation 2.4). The lower it is, the closer the model's predicted values approximates the dependent variable.

$$\text{average prediction error} \equiv \sqrt{\sum_{a=1}^{n}\sum_{b=a+1}^{n}[\text{RIAbi}_b^a - p(\text{RIAbi}_b^a|X_b^a)]^2}. \tag{2.4}$$

Finally the last column lists the value of each model's log likelihood. The lowest prediction error, highest percentage correctly classified and highest log likelihood are each time indicated in bold.

All three statistics show that while the added predictive power of corruption is small when compared to that of the geographic variables, the increase is comparable with that of the other political variables (colonial history and the level of democracy). The explanatory power of both political variables exceeds that of the economic variables (GDP and the capital-labor ratios). In line with the conclusion of the previous section, the additional predictive power of the squared value of corruption is negligible.

Table 2.6: Comparison of the predictive power of the various models

	Model	Table	Column	% Correctly classified[(a)]	Av. Pred. error[(b)]	Log likelihood
1	Distance	2.1	1	75.07	182.53	-206.35
	Economic	2.1	2	75.71	181.56	-197.84
	Economic & political	2.1	7	**84.30**	**177.47**	**-192.65**
2	Economic & political	2.1	7	77.94	159.37	-473.04
	ICRG (average)	2.12	2	**78.66**	150.21	-473.04
	ICRG & $ICRG^2$ (average)	2.3	2	78.56	**150.05**	**-472.70**
3	Economic and political	2.1	7	83.19	48.07	-156.98
	ICRG (separate)	2.13	2	**83.76**	45.29	-150.07
	ICRG and $ICRG^2$ (separate)	2.5	2	83.19	**45.19**	**-149.88**
4	Economic & political	2.1	7	75.66	187.60	-554.11
	WGI (average)	2.12	4	**76.39**	178.48	-554.11
	WGI & WGI^2 (average)	2.3	4	76.12	**178.17**	**-553.44**
	CPI (average)	2.12	6	**78.03**	170.29	-528.21
	CPI & CPI^2 (average)	2.3	6	77.85	**170.26**	**-528.05**
5	Economic & political	2.1	7	77.81	109.14	-341.70
	WGI (separate)	2.13	4	79.66	101.39	-323.73
	WGI & WGI^2 (separate)	2.5	4	**79.80**	**101.41**	**-323.51**
	CPI (separate)	2.13	6	**80.09**	99.06	-314.83
	CPI & CPI^2 (separate)	2.5	6	79.52	**98.75**	**-313.54**

This table provides a one-by-one comparison of the explanatory power of various models. The comparisons are made such that both models operate on the same sample. [(a)]The percentage of observations that are correctly predicted by the model. [(b)]Average prediction error (equation 2.4).

2.7 Conclusion

This chapter explores the motives behind the proliferation of regional integration agreements in Africa. We focus on rent-seizure, rent-destruction and rent-shielding because they are able to explain both the growth of African agreements as well as the lack of progress in liberalizing intra-African trade. Static and dynamic analysis on the other hand predict strong welfare-reducing effects, and most new regionalism theories rely on strong economic integration.

While most theoretical models posit a positive relation between rent-seeking and trade agreements, the underlying mechanism can be starkly different. Broadly speaking, rent-seizure hypotheses argue that a corrupt government will enter into agreements to foster rent-seeking behavior, while rent-shielding and rent-destruction remonstrate that the opposite holds true. All three models predict that at intermediate levels of corruption, an increase in corruption will raise the probability of an agreement. However, rent-destruction and rent-shielding also predict a reversal at high levels of corruption.

We find that corruption does have a significant impact on the willingness to enter a trade agreement. In line with theoretical predictions an increase in the level of corruption raises the probability of there being a trade agreement, especially if this increase happens in the larger of the two countries. This result is robust for the choice of corruption indicator, corrections for endogeneity and whether or not we take into account that intra-African agreements have more than two partner countries. In contrast to the predictions of the rent-destruction and rent-shielding hypotheses, we do not find strong evidence of a reversal at high levels of corruption. This could be seen as very tentative support of the rent-seizing hypothesis of Grossman and Helpman (1995) when explaining intra-African trade agreements, but further research is needed.

While the effect of corruption is significant, the most important factor explaining trade integration is distance. The closer countries lie together and the more remote they are, the higher the probability of an agreement. While the overall explanatory power of rent-seeking behavior is not large, it has similar explanatory power to shared colonial history and even exceeds that of GDPs and capital-labor ratios.

While these results confirm that corruption has had an impact on the formation of African trade agreements, they cannot differentiate between the different motivations for the agreements: i.e. whether they were closed to enable or combat corruption. While it might not be possible to separate them statistically, future research could look at how the level of corruption is subsequently affected by trade agreements. This would answer which theory prevailed ex post, even if it can not gauge the original motives behind the agreements. More importantly, finding out how corruption is affected by integration agreements would have important lessons for development policy.

References

Acemoglu D., Johnson, S.H. and Robinson, J.A. (2001) The colonial origins of comparative development: An empirical investigation. *The American Economic Review* 91(5):1369–1401.

Arcand, J.L., Olarreaga, M. and Zoratto, L. (2011) Weak governments and trade agreements. Centre for Economic Policy Research discussion paper 8595.

Baier, S.L. and Bergstrand, J.H. (2002) On the endogeneity of international trade flows and free trade agreements.

Baier, S.L. and Bergstrand, J.H. (2004) Economic determinants of free trade agreements. *Journal of International Economics* 64(1):29–63.

Cameron, A.C., Gelbach, J.B. and Miller, D.L. (2006) Robust inference with multiway clustering. University of California, Davis, Department of Economics working paper 99.

Carrere, C. (2004) African regional agreements: impact on trade with or without currency unions. *Journal of African Economies* 13(2):199–239.

Decaluwe, B., Njinkeu, D. and Cockburn, J. (1997) A UDEAC case-study. In Oyejide, A., Elbadawi, I. and Yeo, S. (editors) *Regional Integration and Trade Liberalization in SubSaharan Africa, vol.3: Regional Case-Studies.* St. Martin's Press, New York, pp. 86–147.

Endoh, M. (2006) Quality of governance and the formation of preferential trade arrangements. *Review of International Economics* 14, 758–772.

Ethier, W.J. (2011) The political-support view of protection. Penn Institute for Economic Research working paper 11-026.

Feenstra, R.C., Inklaar, R. and Timmer, M.P. (2013) The next generation of the Penn world table.

Foroutan, F. and Pritchett, L. (1993) Intra-Sub-Saharan African trade: is it too little? *Journal of African Economies* 2(1):74–105.

Greene, W.H. (2002) *Econometric Analysis.* Pearson Education, New Jersey.

Grossman, G.M., Helpman, E. (1995) The politics of free-trade agreements. *The American Economic Review* 85(4):667–23.

Iapadre, L. and Luchetti, F. (2010) Trade regionalism and openness in Africa. European University Institute, Robert Schuman Center for Advanced Studies RSCAS 2010/54.

Marshall, M.G., Gurr, T.R. and Jaggers, K. (2014) Polity IV project: political regime characteristics and transitions, 1800-2013. Center for Systemic Peace.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010) The worldwide governance indicators - methodology and analytical issues. The World Bank.

Khandelwal, P. (2004) COMESA and SADC: Prospects and challenges for regional trade integration. IMF Working Papers 04/227.

Krugman, P. and Venables, A.J. (1989) Integration and the competitiveness of the peripheral industry. Centre for Economic Policy Research discussion paper 363.

Landes, D.S. (1998) *The wealth and poverty of nations: why some are so rich and some so poor.* W.W.Norton & Co., New York.

Maggi, G. and Rodríguez-Clare, A. (1998) The value of trade agreements in the presence of political pressures. *Journal of Political Economy* 106(3): 574–601.

Mansfield, E.D., Milner, H.V. and Rosendorff, B.P. (2002) Why democracies cooperate more: Electoral control and international trade agreements. *International Organization* 56(3):477–513.

Márquez-Ramos, L., Martínez-Zarzoso, I. and Suárez-Burguet, C. (2011) Determinants of deep integration: Examining socio-political factors. *Open Economic Review* 22(3):479–500.

Mauro, P. (1995) Corruption and growth. *The Quarterly Journal of Economics* 110(3):681–712.

M'Bet, A. (1997) The CEAO and UEMOA within ECOWAS: The road ahead towards West-African economic integration. In Oyejide, A., Elbadawi, I. and Yeo, S. (editors), *Regional Integration and Trade Liberalization in SubSaharan Africa, vol.3: Regional Case-Studies.* St. Martin's Press, Inc., New York, pp. 66–85.

Newey, W.K. (1987) Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36:231–250.

Nunn, N. (2009) The slave trade and the origins of mistrust in Africa. Natural Bureau for Economic Research working paper 14783.

Ornelas, E. (2005) Rent destruction and the political viability of free trade agreements. *The Quarterly Journal of Economics* 1475–1506.

Sachs, J.D. and Warner, A.M. (1995) Natural resource abundance and economic growth. Natural Bureau for Economic Research working paper 5398.

Schuman, R. (1950) Declaration of 9 may 1950.

Söderbaum, F. (2004) *The Political Economy of Regionalism: The Case of Southern Africa.* Palgrave Macmillan, New York.

The World Bank (2005) Global Economic Prospects: Trade, Regionalism, and Development. The International Bank for Reconstruction and Development, Washington (DC).

Transparency International (2012) CPI index - in detail.

Treisman, D. (2007) What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science* 10:211–244.

UNU-CRIS (2006) Regional integration knowledge system. United Nations University, Comparative Regional Integration Studies.

Viner, J. (1950) *The Customs Union Issue.* Stevens and Sons, London.

Winters, A.L. (2001) Post-lomé trading arrangements: The multilateral alternative. In: von Hagen, J., Widgren, M. (Eds.), *Regionalism in Europe. Geometries and Strategies After 2000.* Kluwer Academic, The Netherlands, Ch. 10, pp. 221–260.

Woolridge, J.M. (2005) Unobserved Heterogeneity and Estimation of Average Partial Effects. *Identification and Inference for Econometric Models.* Cambridge University Press, Ch. 3, pp. 27–55.

Wu, J.P. (2006) Measuring and explaining levels of regional economic integration. In De Lombaerde, P. (editor) *Assessment and Measurement of Regional Integration.* Routledge, New York, Ch. 9, pp. 162–179.

Yang, Y. and Gupta, S. (2007) Regional trade arrangements in Africa: Past performance and the way forward. *African Development Review* 19(3):399–431.

Yeats, A. J. (1998) What can be expected from African regional trade arrangements? World Bank Policy Research working paper 2004.

Appendices

2.A Schematic overview empirical literature

Paper	Framework	Indicator	Monotonic?	Corruption ↑
Wu (2006)	Unilateral	Corruption (CPI)	Yes	Probability RIA ↑
Endoh (2006)	Bilateral	Governance (WGI)	Yes	Probability RIA ↓
Arcand et al. (2011)	Bilateral	Corruption (inferred from tariff data)	Includes squares and interaction with bargaining power	Inverted U pattern and significant interaction terms

2.B First stage regressions

Table 2.7: First stage regressions

	(1) ICRG	(2) WGI	(3) CPI
Tropics	−0.159	−0.236***	−0.153***
	(−1.41)	(−4.34)	(−0.322)
Settler mortality	−35.19*	–	−8.653
	(−1.71)		(7.073)
British colony	–	0.0725*	–
		(1.98)	
Constant	0.704***	0.478***	0.338***
	(6.34)	(9.11)	(7.79)
Observations	28	47	39
R^2	0.214	0.362	0.308
F-test	3.65**	12.95***	9.44***

Linear regression of ICRG, WGI and CPI on different instruments. t-statistics in parentheses. *, **, *** indicates significance at 10%, 5%, and 1% level.

2.C Summary statistics of the dependent variables

Table 2.8: Summary statistics

Variable	Obs.	Mean	St.dev	Min	Max
Natural	1378	-8.036	0.700	-9.188	-1.743
Remote	1378	8.187	0.122	7.914	8.536
Adjacent	1378	0.073	0.260	0	1
Landlocked	1378	0.490	0.500	0	1
GDP_a	1368	9.109	1.347	4.681	12.166
DKL	1128	1.165	0.837	0.002	3.791
DROWKL	1128	0.323	1.150	-2.726	3.370
Colony	1378	0.253	0.435	0	1
$Polity_a$	1346	-5.338	4.374	-10	9.800
$ICRG_a$	1113	0.448	0.197	0	1
WGI_a	1378	0.301	0.139	0	0.644
CPI_a	1378	0.178	0.108	0	0.500

Correlation table

	Natural	Remot.	Adj.	Landl.	GDP_a	DKL	DROWKL	Colony
Natural	1							
Remote	-0.316	1						
Adjacent	0.519	-0.052	1					
Landlocked	0.046	-0.228	0.051	1				
GDP_a	0.042	-0.129	0.064	-0.124	1			
DKL	0.001	-0.024	-0.001	0.010	0.030	1		
DROWKL	0.080	-0.210	0.057	0.268	-0.053	0.170	1	
Colony	0.094	0.044	0.121	0.042	0.077	-0.045	-0.034	1
$Polity_a$	-0.082	0.172	-0.021	0.027	0.202	-0.002	-0.114	0.022
$ICRG_a$	-0.098	0.290	0.012	0.018	0.103	0.047	-0.219	0.050
WGI_a	-0.180	0.484	-0.052	-0.030	0.049	0.078	-0.147	0.068
CPI_a	-0.161	0.389	-0.035	-0.089	0.104	0.109	-0.239	0.044

	$Polity_a$	$ICRG_a$	WGI_a	CPI_a
$Polity_a$	1			
$ICRG_a$	0.280	1		
WGI_a	0.332	0.553	1	
CPI_a	0.370	0.498	0.800	1

2.D Unilateral regressions

Table 2.9: Unilateral regressions

	ICRG		WGI	
	(1)	(2)	(3)	(4)
Landlocked	–	−0.185	–	−0.832*
		(−0.50)		(−1.74)
Land area	–	0.000179	–	0.000213
		(0.67)		(0.60)
Island	–	0.336	–	−0.478
		(0.36)		(−0.52)
GDP	0.00184		0.0701	
	(0.08)		(0.37)	
GDP/cap	–	−0.0515	–	−0.148
		(−0.77)		(−1.59)
Population	–	0.00169	–	−0.00165
		(0.21)		(−0.17)
Polity	–	−0.00759	–	−0.0191
		(−0.26)		(−0.44)
Corruption	−0.587	0.667	−21.47	1.667
	(−0.14)	(0.19)	(−1.22)	(0.16)
Corruption^2	3.262	0.820	33.80	−2.503
	(0.64)	(0.19)	(1.46)	(−0.19)
GDP × Corruption	−0.0440	−0.00999	−0.691	0.00319
	(−0.81)	(−0.62)	(−1.13)	(0.16)
LR$^{(a)}$ corr	0.535	0.706	0.310	0.995
Constant	–	−1.181	–	−0.806
		(−1.40)		(−0.38)
Loglikelihood	−83.88	−128.4	−31.22	−80.51
Observations	210	209	99	149
n^o of countries	36	36	33	50

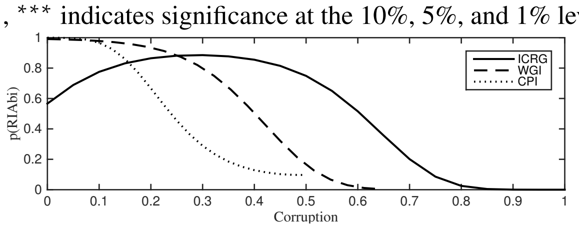
Columns 1 and 3 are fixed effects and 2 and 4 are random effects logistic regressions of RIAuni on corruption and controlling variables. t-statistic in parentheses. $^{(a)}$p-value of the likelihood ratio test on the joint significance of the corruption variables. * indicates significance at 10% level.

2.E IVprobit regressions

Table 2.10: Average-levels ivprobit regressions

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.536***	1.409***	1.380***	1.329***	1.481***	1.366***
	(14.73)	(12.40)	(13.66)	(13.65)	(13.20)	(12.48)
Remote	5.439***	5.548***	6.217***	6.463***	7.515***	6.711***
	(8.35)	(7.12)	(8.68)	(8.16)	(7.52)	(9.20)
Adjacent	–	1.110***	–	0.866**	–	0.769**
		(3.02)		(2.45)		(2.17)
Landlocked	–	0.291**	–	0.183	–	0.157
		(2.05)		(1.58)		(1.40)
GDP_{av}	−0.00348	−0.00850	−0.0271	−0.0374	−0.0629	−0.102
	(−0.04)	(−0.09)	(−0.38)	(−0.52)	(−0.86)	(−1.34)
GDP_{diff}	0.0988	0.139**	0.0454	0.0711	0.110*	0.114**
	(1.56)	(2.08)	(0.87)	(1.36)	(1.82)	(1.99)
DKL	0.0690	0.0979	0.0347	0.0612	0.0958	0.126**
	(0.92)	(1.26)	(0.63)	(1.11)	(1.50)	(2.03)
DROWKL	−0.135**	−0.177***	−0.0563	−0.101**	−0.254***	−0.239***
	(−2.50)	(−2.72)	(−1.12)	(−2.32)	(−4.39)	(−4.58)
Colony	–	0.228*	–	0.427***	–	0.359***
		(1.71)		(3.66)		(3.13)
$Polity_{av}$	–	$-5.54e-4$	–	0.0227	–	0.0217
		(−0.02)		(1.27)		(0.60)
$Corruption_{av}$	4.445	7.112	−17.84	−2.693	−7.249	−20.14
	(0.68)	(1.09)	(−0.96)	(−0.24)	(−0.31)	(−1.04)
$Corruption^2_{av}$	−9.005	−12.21**	17.13	−8.000	−13.66	20.53
	(−1.48)	(−2.06)	(0.60)	(−0.47)	(−0.23)	(0.44)
Wald$^{(a)}$ corr	$1.08e-5$	$1.87e-6$	$1.38e-7$	$6.97e-10$	$1.18e-09$	$7.07e-10$
Wald$^{(a)}$ exog	$1.57e-3$	$6.12e-5$	$6.99e-3$	$3.4e-4$	0.01	0.033
Constant	−32.91***	−35.90***	−36.21***	−40.96***	−48.10***	−41.26***
	(−6.89)	(−6.47)	(−6.35)	(−7.22)	(−5.25)	(−7.48)
Observations	970	970	1125	1124	1097	1097
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

ivprobit regression of RIAbi on corruption and control variables. The instruments are interacted in the same way as the corruption variables. t-statistics in parenthesis. $^{(a)}$p-value of the Wald test of the joint significance of the corruption values and their exogeneity (H_0: corruption is exogenous). *, **, *** indicates significance at the 10%, 5%, and 1% level.

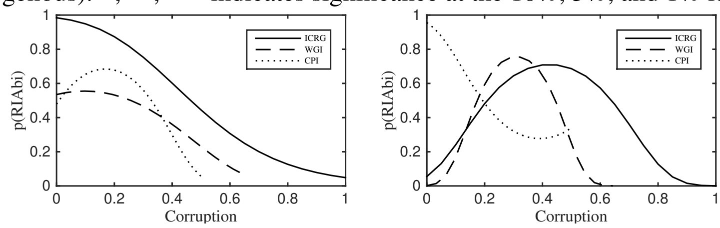


The marginal effect of average corruption on the probability of joining a RIA in the average-levels ivprobit regressions, keeping all other variables at their mean values. The maximum values of WGI and CPI in Africa are 0.62 and 0.5, while ICRG's values range between 0 and 1.

Table 2.11: Individual-levels ivprobit regression

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.731***	1.592***	1.533***	1.425***	1.845***	1.779***
	(9.00)	(7.29)	(15.24)	(12.00)	(12.36)	(11.44)
Remote	8.567***	9.082***	6.674***	4.975*	7.439***	6.983***
	(5.23)	(4.32)	(6.05)	(1.79)	(6.28)	(5.29)
Adjacent	–	1.156*	–	0.673*	–	0.228
		(1.82)		(1.65)		(0.57)
Landlocked	–	0.280	–	−0.0310	–	−0.00452
		(0.62)		(−0.10)		(−0.02)
GDP$_a$	0.190	0.220	0.0813	0.0649	0.137	0.109
	(1.49)	(1.46)	(1.29)	(0.92)	(1.51)	(1.50)
GDP$_b$	−0.0788	−0.137	0.0562	−0.0427	−0.137*	−0.111
	(−0.43)	(−0.72)	(0.93)	(−0.35)	(−1.65)	(−1.31)
DKL	0.0425	0.0569	0.0426	0.0754	0.0724	0.0745
	(0.35)	(0.43)	(0.67)	(0.96)	(0.85)	(0.89)
DROWKL	0.112	0.0293	0.0607	−0.0734	0.0285	0.0329
	(0.71)	(0.14)	(0.81)	(−0.62)	(0.35)	(0.42)
Colony	–	0.107	–	0.160	–	0.131
		(0.40)		(0.52)		(0.79)
Polity$_a$	–	0.00382	–	−0.00538	–	−0.00912
		(0.10)		(−0.35)		(−0.33)
Polity$b$	–	−0.00572	–	0.0757**	–	0.0258
		(−0.14)		(2.10)		(1.05)
Corruption$_a$	−2.827	−5.352	−0.271	1.047	10.99	6.338
	(−0.44)	(−0.71)	(−0.04)	(0.11)	(1.02)	(0.51)
Corruption^2_a	−0.0270	1.542	−4.673	−5.396	−29.42	−18.83
	(−0.01)	(0.31)	(−0.55)	(−0.50)	(−1.38)	(−0.80)
Corruption$_b$	6.153	10.16	−3.472	23.34	−17.27***	−11.77
	(1.01)	(1.40)	(−0.46)	(1.13)	(−2.58)	(−1.32)
Corruption^2_b	−7.044	−11.98	1.248	−37.73	27.32*	15.05
	(−0.99)	(−1.36)	(0.12)	(−1.35)	(1.89)	(0.78)
Wald$^{(a)}$ corr$_a$	0.0998	0.180	$5.01e-6$	0.0399	0.0178	0.0300
Wald$^{(a)}$ corr$_b$	0.600	0.371	$1.75e-4$	$7.17e-6$	0.00886	0.0375
Wald$^{(a)}$ exog	0.557	0.400	0.0895	0.0171	0.0856	0.388
Constant	−57.19***	−62.44***	−41.98***	−31.94	−44.85***	−41.69***
	(−5.18)	(−4.09)	(−5.46)	(−1.61)	(−4.78)	(−4.24)
Observations	351	351	990	946	703	703
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

ivprobit regression of RIAbi on corruption and control variables. $^{(a)}$p-value of the Wald test of the joint significance of the corruption values of each country and their exogeneity (H_0: corruption is exogenous). *, **, *** indicates significance at the 10%, 5%, and 1% level.



Country A Country B

Surface plots of the marginal effect corruption and GDP on the probability of joining a RIA in the individual-levels ivprobit regressions, keeping all other variables at their mean values.

2.F Regressions without corruption squared

Table 2.12: Average-levels without corruption squared

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.508***	1.409***	1.409***	1.313***	1.476***	1.388***
	(11.68)	(8.98)	(11.78)	(10.83)	(14.57)	(12.29)
Remote	5.480***	6.118***	5.839***	6.396***	7.015***	6.429***
	(5.25)	(4.62)	(5.32)	(3.53)	(4.06)	(3.18)
Adjacent	–	0.914***	–	0.803***	–	0.832***
		(9.70)		(9.81)		(9.22)
Landlocked	–	0.364**	–	0.197	–	0.128
		(2.44)		(0.98)		(0.65)
GDP_{av}	−0.0584	−0.0474	−0.0605	−0.0950	−0.0823	−0.127
	(−0.85)	(−0.54)	(−0.75)	(−1.00)	(−0.95)	(−1.35)
GDP_{diff}	0.105**	0.146***	0.0838**	0.116***	0.113***	0.135***
	(2.32)	(3.19)	(2.08)	(3.60)	(3.21)	(4.92)
DKL	0.00520	0.00808	0.0370**	0.0584**	0.0925***	0.109***
	(0.39)	(0.46)	(2.22)	(2.22)	(3.56)	(3.04)
DROWKL	−0.141*	−0.215***	−0.0701	−0.0924***	−0.231***	−0.226***
	(−1.87)	(−3.98)	(−1.53)	(−4.17)	(−6.38)	(−6.45)
Colony	–	0.369***	–	0.455***	–	0.342***
		(7.31)		(5.33)		(4.25)
Polity_{av}	–	−0.0264***	–	0.0178	–	0.0316**
		(−2.99)		(1.23)		(2.57)
Corruption_{av}	−4.949***	−6.427***	−5.966***	−7.783***	−11.46***	−10.83***
	(−4.21)	(−5.68)	(−5.35)	(−5.18)	(−5.22)	(−4.09)
Error term ε	3.515***	4.991***	3.065**	5.139***	4.530**	3.698
	(2.71)	(4.22)	(2.45)	(4.17)	(2.54)	(1.58)
Constant	−30.81***	−37.08***	−34.80***	−39.73***	−43.67***	−39.55**
	(−3.77)	(−3.58)	(−3.95)	(−2.74)	(−3.21)	(−2.46)
Observations	970	970	1125	1124	1097	1097
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

Probit regression of RIAbi on corruption, the error term from the first stage regressions (ε) and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. *, **, *** indicates significance at the 10%, 5% and 1% level.

Table 2.13: Indiviual-levels probit regression without corruption squared

	ICRG		WGI		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Natural	1.791***	1.675***	1.554***	1.420***	1.844***	1.800***
	(9.66)	(9.58)	(10.99)	(10.20)	(10.01)	(9.77)
Remote	8.696***	8.385***	6.360***	6.861***	8.904***	8.117***
	(5.16)	(4.35)	(4.77)	(3.63)	(4.28)	(3.43)
Adjacent	–	0.919***	–	0.541***	–	0.176*
		(12.74)		(6.52)		(1.75)
Landlocked	–	−0.218	–	0.165	–	0.0541
		(−1.20)		(0.87)		(0.53)
GDP_a	0.155***	0.166***	0.0623***	0.0916***	0.0378	0.0918***
	(5.94)	(4.00)	(4.11)	(3.78)	(1.00)	(3.10)
GDP_b	0.0149	−0.0683	0.0361	0.0512	−0.00796	−0.0266
	(0.07)	(−0.45)	(0.84)	(1.08)	(−0.19)	(−0.50)
DKL	0.0187	0.0545*	0.0340	0.0419	0.0847***	0.103***
	(0.71)	(1.69)	(0.94)	(1.23)	(4.88)	(7.32)
DROWKL	0.125	0.147	0.0754	0.0810***	−0.0268	−0.0183
	(0.56)	(0.63)	(1.55)	(2.75)	(−0.46)	(−0.37)
Colony	–	0.0147	–	0.439***	–	0.214*
		(0.20)		(4.86)		(1.89)
$Polity_a$	–	0.00594	–	−0.00996	–	−0.0258***
		(0.41)		(−1.61)		(−2.72)
Polity*b*		0.0181		0.0262*		0.0407**
		(0.38)		(1.91)		(2.31)
$Corruption_a$	−2.991*	−3.114*	−3.173***	−3.808***	−4.304***	−4.236***
	(−1.86)	(−1.82)	(−4.16)	(−3.34)	(−3.31)	(−2.83)
$Corruption_b$	0.693	1.350	−2.113*	−3.383***	−8.114***	−6.990***
	(0.34)	(0.63)	(−1.95)	(−5.24)	(−7.77)	(−5.98)
Wald$^{(a)}$ corr	$5.94e-4$	$1.99e-4$	$2.80e-9$	$9.75e-10$	$3.0e-14$	$3.67e-10$
Error term ε_a	1.441	1.697	1.729***	2.563***	−0.174	−0.174
	(0.90)	(1.04)	(3.77)	(3.60)	(−0.19)	(−0.16)
Error term ε_b	−0.602	−1.128	0.167	0.920	4.759***	3.082***
	(−0.29)	(−0.53)	(0.17)	(1.41)	(4.20)	(2.78)
Constant	−57.32***	−55.20***	−38.83***	−44.02***	−56.04***	−50.58***
	(−4.84)	(−3.94)	(−3.72)	(−2.95)	(−3.49)	(−2.75)
Observations	351	351	990	946	703	703
Instruments:	Tropics		Tropics		Tropics	
	Settler mortality		British colony		Settler mortality	

Probit regression of RIAbi on corruption, the error terms from the first stage regressions (ε_a and ε_b) and control variables. t-statistics (in parenthesis) are corrected for multiple non-nested clusters. $^{(a)}$p-value of the Wald tests on the joint significance of the corruption variables. *, **, *** indicates significance at the 10%, 5% and 1% level.

3 | Divining the level of corruption - A Bayesian state-space approach

Abstract

This chapter outlines a new methodological framework for combining indicators of corruption. The state-space framework extends the methodology of the Worldwide Governance Indicators (WGI) to fully make use of the time-structure present in corruption data. It is estimated using a Bayesian Gibbs sampler algorithm. The state-space framework holds many advantages from a practical, an estimation and a theoretical point of view. Most importantly, it significantly expands the period for which the index can be computed while at the same time addressing the selection bias issues that trouble the Corruption Perceptions Index (CPI). In addition, its estimates are more stable and have smaller confidence intervals than both CPI and WGI. Finally, the estimation procedure is explicit in its assumptions and allows the data to be entered without any ex-ante imputations resulting in an index that is less affected by ad hoc modeling choices.

Keywords: Corruption perception; State-space model; Bayesian econometrics; Worldwide Governance Indicators.

JEL classification: C43; O17; O57; P16; P26.

3.1 Introduction

Researchers looking at the effects or determinants of corruption are faced with the difficulty of having to choose one out of the more than 70 individual indicators available. Each indicator differs in availability in time, countries covered, exactly what it is trying to measure, and where or with whom it was collected. Because that one indicator that meets all requirements often proves elusive, most studies resort to aggregated indicators of corruption. The two most used are the Corruption Perceptions Index (CPI) published by Transparency International and the Worldwide Governance Indicators' index of corruption (WGI) made available by the World Bank.

However, the use of these aggregated indicators is not without criticism, especially when making comparisons over time. Shifts in the indices are not only driven by the level of corruption, but by changes in the methodology and sources as well. Moreover, both indicators only go back to the mid-nineties and early values suffer from serious selection bias problems (Treisman, 2007).

This chapter outlines a new methodology for combining indicators of corruption that fixes a number of the issues plaguing the WGI and CPI. Starting from the WGI methodology, the state-space model uses the persistence of corruption to better identify actual changes in the level of corruption. This leads to smaller confidence intervals, especially when only a few indicators are available.

Following Høyland et al. (2012), the model is estimated using a Bayesian Gibbs sampling algorithm. In combination with the solution to missing data, this allows the model to be estimated without additional assumptions or manipulations to the data (imputations, sub-level aggregations, etc.). Moreover, the flexibility of the Gibbs sampling algorithm makes it easy to extend the model to allow cross-correlated or persistent measurement errors in the individual indicators. Furthermore, it can be combined with the approach of Givens (2013) and capture multiple aspects of governance in the style of an exploratory principal component analysis. The resulting indicator is dubbed the Bayesian Corruption Indicator and is available for download at `www.sherppa.be`.

The following section reviews the data and methodology used in the CPI and the WGI and highlights some of their shortcomings. This is followed by the presentation of the new framework after which its results are discussed.

3.2 Individual indicators of corruption

An important question when choosing the sources for an indicator of corruption is whether to use incidence or perception-based surveys. The former asks for personal experience with corruption within the last x months, while the latter asks respondents for their opinion on the level of corruption in the country as a whole, or in various branches of the government. Following the CPI and the WGI, this chapter focuses explicitly on corruption perception[1]. For comparability's sake, the same sources as the WGI are used, which include firm and household survey data as well as expert assessments from governmental, NGO and commercial institutions (table 3.1).

Table 3.1: Overview of perceived corruption indicators

Cross-country survey of households	**Expert assessment from commercial risk rating agencies**
- Gallup World Poll	- International Country Risk Guide
- Latinobarometer	- Global Risk Service
	- World markets online
Expert assessment from NGO and think tanks	- Economist Intelligence Unit
- Global Integrity	- Political and Economic Risk Consultancy
- The Freedom House	
- Bertelsmann Transformation Index	**Expert assessment from governments and multilaterals**
- Global Corruption Barometer	- Country Policy and Institutional Assessment
	- African Development Bank
Cross-country survey of firms	- Asian Development Bank
- Afrobarometer	- IFAD Rural Sector Performance Assessment
- Global competitiveness survey	- Institutional Profiles database
- Vanderbilt University's Americas Barometer	
- World Competitiveness Yearbook	
- Business Environment and Enterprise Performance Survey	

(Arndt and Oman, 2006)

[1]For a discussion on the merits and demerits of perception versus experience based measures of corruption, see e.g. Kaufmann et al. (2004); Lambsdorff (2005); Kaufmann et al. (2007a,b); Treisman (2007); Kaufmann et al. (2010) or Roca (2011).

However, there are a few differences with the WGI dataset. First of all, whenever possible, the individual indicators/survey questions are used instead of sub-level aggregations. While this leads to shorter time-series for some indicators, it also minimizes measurement errors due to changes in the composition of the grouped indicator. The effect of using the individual indicators on the overall precision of the index is discussed in more detail in section 3.7.3.

The individual survey questions were separated into indicators measuring corruption perception, experience and anti-corruption measures. For example, a survey question asking for '*personal experience with corruption during the last year*' (Global Corruption Barometer) is considered to measure experience and is left out. On the other hand, a more broad question about the experience with corruption of '*firms in my line of business*' (Business Environment and Enterprise Survey) is deemed to measure perception.

In contrast to the WGI, indicators of the (perception of) the efficiency of anti-corruption measures were also left out. Anti-corruption measures capture a cause of corruption and the relation between the two is not necessarily positive or linear. A country with high levels of corruption that it is actively trying to fight would have high corruption and strong measures. A country without corruption might have no need for strong anti-corruption institutions or on the contrary have little corruption because of strong institutions. In addition, removing these indicators makes it possible to test whether institutions are successful in bringing down the perceived level of corruption.

A final distinction between the two datasets is that indicators available every couple of years (e.g. the Latinobarometer) were only used in one year. The WGI on the other hand impute the data onto all intervening years as well. How the state-space model subsequently deals with these missing observations is described in more detail in section 3.4.2.

Constructed in this way, the corruption perception dataset contains 69 variables coming from 18 different sources and covers 211 countries and regions. While the dataset starts in 1984, ICRG is the only available indicator for the first 10 years.

From then on the number of available indicators rapidly increases and remains high up to 2012. A full description of the sources and their individual indicators of corruption can be found in 3.B. For a more thorough analysis of the sources see Arndt and Oman (2006, p. 52-57).

3.3 Composite indicators of corruption

3.3.1 Corruption Perceptions Index

Published yearly since 1995 by Transparency International, the Corruptions Perceptions Index (CPI) is probably the most widely known composite indicator of corruption. As the name indicates, it combines perception-based indicators in order to capture the *'misuse of public power for private benefit'* (Lambsdorff, 2005, p.4). The higher a country's CPI score, the lower its level of corruption.

It merges the indicators by first standardizing their values and then taking a simple average. Up until 2012, they used a technique called *matching percentiles* to standardize the *rank* data from indicators. However, since 2012 they have changed this to a normalization of the nominal values of the indicators, which is then adjusted to ensure backwards compatibility. If y is the original indicator value, its normalized equivalent $y^\star$ is computed as:

$$y^\star = \frac{y - mean(y)}{std(y)} * sign * 20 + 45. \tag{3.1}$$

The sign variable ensures that all variables associate an increase in the index with a drop in the level of corruption. These standardized values are rescaled and capped to lie between 0 and 100 (Saisana and Saltelli, 2012).

The methodology used in creating the CPI index has a number of drawbacks, the most important of which is that it should not be used for comparisons over time (Transparency International, 2012). Because the matching percentiles technique uses only the rank data and combines it with a simple average, the index is very sensitive to changes in countries covered and indices used. While the change in methodology in 2012 has improved on this, the pre-2012 values are still computed using the old methodology.

Secondly, it uses only a subset of the available corruption indicators and does not include countries for which there are less than three sources available in a given year. As a result, the coverage of the indicator is limited, especially for its earlier values. Moreover, the selection is not independent of the level of corruption, causing the index to be prone to a selection bias issue (Treisman, 2007). In order to alleviate some of the availability problems, the data is manipulated in a number of ad-hoc ways. For example, some (but not all) sources are averaged over the last three years, while others are used twice.

Lastly, there is the issue of apparent randomness in the weighing of the individual indicators. The updated methodology makes the arbitrary choice of multiplying with 45, adding 20 and capping all values over 100 purely for reasons of continuity. However, there are no clear ex ante reasons why the variables should be rescaled in this way and changing these weights would lead to different values and rankings that are equally valid.

3.3.2 Worldwide Governance Indicators

The Worldwide Governance Indicators' index of corruption measures *'the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as 'capture' of the state by elites and private interests.* (Kaufmann et al., 2010, p.4). To do this, they combine the perception based corruption indicators using an unobserved components (or factor) model.

Unobserved components model

Denoting the $(k \times 1)$ vector of corruption indicators with $y_{i,t}$, and the unknown 'true' level of corruption with the scalar $\alpha_{i,t}$, they estimate the following model:

$$y_{i,t} = C_t + Z_t \alpha_{i,t} + \varepsilon_{i,t} \quad (3.2)$$

$$\alpha_{i,t} \sim N(0,1) \quad (3.3)$$

$$\varepsilon_{i,t} \sim N(0,H_t) \quad (3.4)$$

for all countries $i = 1,\ldots,p$ and years $t = 1,\ldots,n$, with k the number of individual corruption indicators in $y_{i,t}$.

Re-estimated in each year, the $(k \times 1)$ vectors C_t and Z_t capture both differences in scaling as well as the distribution of values. For example, some indicators might easily assign the highest score, while others reserve that only for a limited numbers of countries. The unknown level of corruption α is assumed to be normally distributed.[2] To identify this model, the choice of units of α in each year is fixed at mean zero and standard deviation of one (equation 3.3). Kaufmann et al. (2010) argue that this choice of units does not preclude their use in time-series or panel studies because they find no significant evidence of a worldwide trend in corruption.

Finally, $\varepsilon_{i,t}$ is an error term with $(k \times k)$ variance matrix, H_t. The measurement errors of different indicators are assumed to be uncorrelated: $E(\varepsilon'_{i,t}\varepsilon_{j,t}) = 0 \; \forall i \neq j$, which means that H is a diagonal matrix. The error term is meant to capture two effects. Firstly, it will account for errors in the data collection process. Secondly, it also corrects for the possibility that the indicators do not measure the overall level of corruption, but a related concept like the level of petty corruption, or the level of corruption in the judiciary.

[2] Høyland et al. (2012) work out a model that relaxes this assumption as it is clear that it is incompatible with the distribution of many of the underlying indicators in y.

Estimation

In order to estimate this model, $\alpha_{i,t}$ and $\varepsilon_{i,t}$ are assumed to be multivariate normal distributed. The data then is split up in a *representative* and *non-representative* group. Simply put, the representative group contains all indicators whose scope either covers the entire population, or represents a random selection of countries. Conversely, the non-representative group contains those indicators whose coverage is not independent from the level of corruption. For example the Freedom House index, which focuses on Eastern European countries.

In the first step, estimation is done using only the representative group, where the yearly expected value of α is assumed zero. These estimates are then updated with the information from the non-representative group. The advantage of this two-step procedure is that the results from the representative group can be used to assess and correct the bias in the non-representative group without having to make any prior assumptions on the size or direction of the bias.

A final rescaling of the variables ensures that the mean is zero and standard deviation is one in each year. However, selection bias issues require a second rescaling for the earliest index values. Countries that are added later to the sample are found to have lower levels of corruption, meaning that the earlier WGI values are skewed upwards. To compensate, the mean value in each year is adjusted using the values of 2003 as a benchmark.

WGI's index of corruption

The inclusion of the error term with indicator-specific variance is a big advantage of the WGI. It makes it more robust to the inclusion of indicators that are less correlated with the general level of corruption, whether due to measurement errors or because it only measures a related concept (for example, the level of corruption in elected officials). The CPI on the other hand treats all indicators the same, regardless of their reliability or conceptual suitability.

As was the case with the CPI, the initial years of the index are available for a select

Table 3.2: Correlation of the corruption indicators with their lagged values

y^{κ}	$\rho(y_t^{\kappa}, y_{t-1}^{\kappa})$	y^{κ}	$\rho(y_t^{\kappa}, y_{t-1}^{\kappa})$	y^{κ}	$\rho(y_t^{\kappa}, y_{t-1}^{\kappa})$
WGI	0.988	GCB2	0.870	GCS4	0.970
CPI	0.989	GCB27	0.700	GCS5	0.978
ADB	0.905	GCB28	0.575	GCS6	0.973
ASD	0.878	GCB29	0.857	GCS7	0.982
EIU	0.975	GCB3	0.955	GWP	0.931
FRH	0.994	GCB4	0.650	IFD	0.873
GCB1	0.858	GCB5	0.723	LBO1	0.312
GCB10	0.887	GCB6	0.712	LBO3	0.175
GCB11	0.903	GCB7	0.914	LBO4	0.797
GCB12	0.923	GCB8	0.727	PIA	0.951
GCB13	0.919	GCB9	0.782	PRC	0.963
GCB14	0.896	GCS1	0.976	PRS	0.966
GCB15	0.917	GCS2	0.981	WCY	0.976
GCB16	0.971	GCS3	0.974	WMO	0.972

group of countries. The problem is less severe because the index is composed even when only one datasource is available. Nevertheless, from its start in 1996 to 2002 the index is only available every two years. Those values of the individual indicators in the years in between are imputed onto the following year.

3.3.3 Persistence

Because the level of corruption is in a large part driven by social norms and values, it is expected to show a high degree of persistence. Table 3.2 confirms this. While the panel unit root test always rejects the hypothesis that the observed measures of corruption have a unit root for *all* countries, the simple correlation coefficient between the corruption indicator and their lagged values is still greater than 0.9 for more than half of the indicators. In other words, regardless of the rejection of the unit root, the persistence in the observed level of corruption for many countries and indicators is notable.

While this time-dependence is reflected in the values of the CPI and the WGI (table 3.2), they do not make use of it in their estimations. However, by taking the past values into account, the reliability of the estimates can be significantly increased and random ‘noise’ can be better filtered out from the corruption indicators. It will also expand the time period for which the level of corruption can be computed.

In Governance Matters IV, Kaufmann et al. (2005) explore the use of the time-dimension to better assess the level of governance. Using a two-period model, they find that while the overall correlation between the static and dynamic approaches is high, the behavior of the index does change when compared over time. The approach presented in the next section generalizes their model and introduces a different estimation method that simplifies a number of the assumptions used in the WGI model.

3.4 The updated framework

3.4.1 Model

Extending the WGI framework to take the time dependence into account leads to the following system of equations:

$$y_{i,t} = C + Z\alpha_{i,t} + \varepsilon_{i,t} \tag{3.5}$$

$$\alpha_{i,t} = T_i \alpha_{i,t-1} + \nu_{i,t} \tag{3.6}$$

$$\varepsilon_{i,t} \sim N(0, H) \tag{3.7}$$

$$\nu_{i,t} \sim N(0, Q) \tag{3.8}$$

The measurement equation

As before, the measurement equation (3.5) states that the k indicators of corruption $y_{i,t}$ try to measure the 'true' level of corruption $\alpha_{i,t}$. The variables i and t respectively represent the different countries and time-periods. Corruption is defined in the same way as in the Worldwide Governance Indicators (Kaufmann et al., 2010). The scaling parameters C and Z can vary over the indicators of corruption but are kept constant over time and country. This differs from the WGI methodology where these values are recomputed for each year. In doing so, we assume that the relation between the overall level of corruption and a specific question asked using a specific methodology does not change over time.[3] To the extent that the parameters remain constant, reestimating them in each period would imply a significant loss of information. The validity of this assumption is tested in section 3.7.1.

Similarly, the variance of the error term ε can differ over all corruption indicators, but is kept constant over time. Initially cross-correlation between the error terms of different indicators is also ruled out (H diagonal): $E[\varepsilon^{(k)}, \varepsilon^{(m)}] = 0, \forall\, k \neq m$. This assumption is subsequently relaxed to allow the indicators coming from the same source to be correlated (H block-diagonal).

The state equation

The state equation (3.6) allows for the 'true' level of corruption to depend on its previous values. If the values for T_i are set to zero this model coincides with that of the WGI index (equation 3.3). This persistence is modeled as an AR(1) process, restricting the values of T to lie inside the [-1,1] interval. While more complex ways of modeling the persistence might be possible and even desirable, the number of observations per country is too limited to include richer dynamics without having to impose strong a-priori assumptions.

Unlike the parameters of the measurement equation, each country is allowed a different level of persistence, enabling some countries' level of corruption to change more rapidly than that of others.[4] The fact that the unit root tests reject an overall unit root despite the high correlation in the individual indicators (table 3.2) suggests that this level of dependence is to some extent country-specific. Nevertheless, section 3.7.2 discusses the effect of imposing $T_i = \tau$ and $T_i = 1$ as robustness checks.

By bringing the time dimension into play, a lot more information is used in the estimation of each corruption value. Figure 3.1 illustrates this graphically. In the WGI framework, only the current information on corruption, y_t, is used (step b). In

[3] Since the WGI groups indicators per source, this assumption would still imply that its parameters need to be reestimated in each year as the composition of questions per source does not remain constant over time.

[4] While the speed of adjustment is determined both by the value of T and the variance of the innovation in corruption, the latter has been normalized to one to ensure that the model is identified.

the new framework the level of corruption is predicted using both past and future values (step a), after which the information in y_t is used to update that estimate (step b). The importance of step a versus step b will depend on how reliable the corruption indicators (H) are versus how reliable the past values are (T_i and Q).

Because α_{t-1} and α_{t+1} in turn also depend on their past and future values, all available information will be used to estimate the current level of corruption. Not only does this increase the reliability of each estimate, it also helps smooth out the estimates of corruption. By taking the past and future values of corruption into account, the algorithm is better able to distinguish between actual changes in the level of corruption and random measurement errors.

Following Givens (2013), this model could also be used to extract multiple states from the governance indicators, simply by defining $\alpha_{i,t}$ as a vector, rather than a scalar and extracting multiple states. Under the assumptions that each variable is equally reliable ($H = \sigma I$) and disregarding the time-dependence ($T_i = 0$), the model would return a similar exploratory principal component analysis.

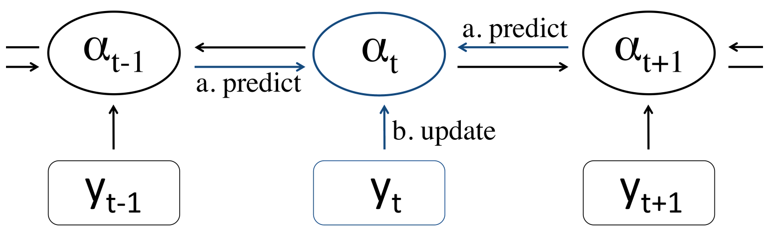


Figure 3.1: Estimation using time dependency

3.4.2 Estimation

This section aims to provide only a very general overview of the estimation technique. More information can be found in 3.A, but for a complete overview of state-space models and how to estimate them, see Kim and Nelson (1999) or Durbin and Koopman (2012).

Gibbs sampling

The updated framework is estimated in a Bayesian framework because of the convenience the Gibbs sampling algorithm provides. Solving the model entails finding the optimal solution for the parameters of the state and measurement equation (C, Z, T and H) as well as the level of corruption (α). While it is possible to solve this model using maximum likelihood, the problem quickly becomes very complex as more and more countries are added. However, using the Gibbs sampling we can split the estimation up into various subcomponents which can be dealt with one at a time. In addition, the modular approach makes it easy to change parts of the estimation model.

Simply put, the Gibbs sampler allows us to draw from a multivariate probability, $p(a,b)$, using only conditional probabilities, $p(a|b)$ and $p(b|a)$. Starting from a certain value b_1, draws are taken iteratively from both conditional distributions while conditioning on the last drawn values:

$$a_1 \sim p(a|b_1) \rightarrow b_2 \sim p(b|a_1) \rightarrow a_2 \sim p(a|b_2) \rightarrow \ldots$$

It can be shown that after a sufficient number of iterations (the burn-in), a_n and b_n represents random draws from the unconditional probability function $p(a,b)$. Using enough random draws, we can then reconstitute the original multivariate probability $p(a,b)$.

In this case, the Gibbs sampler consists of two main components (figure 3.2). In part A, the parameters of the state and measurement equations are drawn conditional on the values for α. Part B samples from the distribution of the 'true' level of corruption while conditioning on the parameters of the state and measurement equation.

An additional advantage of the Gibbs sampling algorithm is that it avoids the need to distinguish between representative and non-representative sources, nor does it require the assumption that α and ε are multivariate normal. The reason is that estimation no longer requires the assumption that the expected value of α is zero.

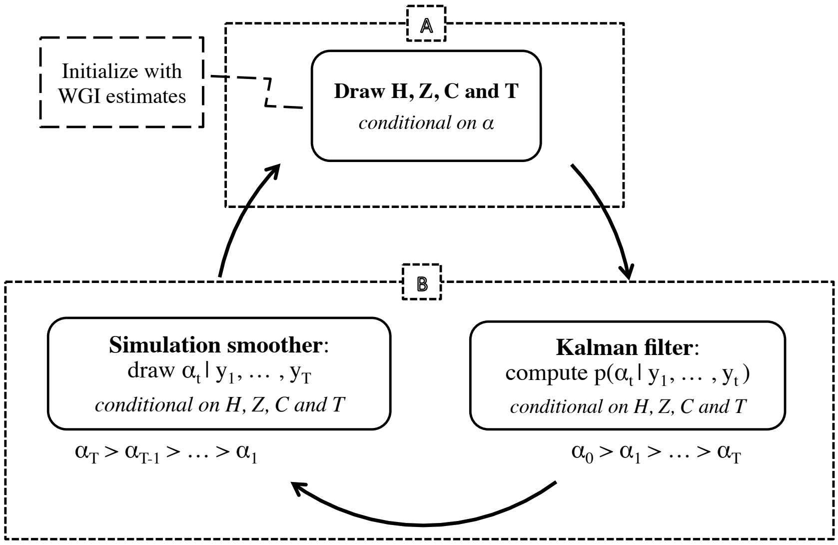


Figure 3.2: Estimation flow chart

If for example an indicator only covers more corrupt countries, it will only be used to update the corruption estimates of those countries it covers (step B, figure 3.2). Similarly, when estimating its scaling parameters (step A), only the information of the relatively more corrupt countries will be used.

More information on the estimation procedure can be found in the appendix 3.A, which also discusses the convergence of the Gibbs sampler. For more details on Bayesian econometrics and Gibbs sampling algorithms, see Lancaster (2004) and Koop et al. (2007).

Missing observations and non-representative indicators

Finally, there is the issue of missing observations. There are different ways of dealing with this in the state-space framework, but they all boil down to the same idea: missing data is replaced by information which is entirely uncertain and consequently holds no value: $y_{missing} = 0$, $Var(\varepsilon_{missing}) = \infty$. As a result, the state-space model will recognize that any information contained in $y_{missing}$ should have no effect on the resulting indicator. This allows the model to run uninterruptedly without fundamentally changing the nature of missing data. In combination with the time dependency, it enables us to significantly increase the number of countries and years for which the index can be calculated without having to impute, group or otherwise manipulate the data (Kim and Nelson, 1999; Durbin and Koopman, 2012).

This solution to missing data also solves the problem of what to do with non-representative indicators: e.g. indicators that only cover the more corrupt countries. To reiterate, the Gibbs sampling algorithm splits up the estimation into two stages. In the first stage the level of corruption is computed and drawn from. Because missing data is replaced by random noise, all the indicators can be included regardless of their availability or selection issues. Moreover, this step is conditional on the parameters C and Z, through which the model will take into account that an index might be unrepresentative. A high value of C will reflect that the indicator covers relatively more corrupt countries (leading a lower estimated BCI index) and vice versa.

In the second stage of the Gibbs sampler the parameters are computed conditional on the value of corruption. Say we have an indicator that only covers more corrupt countries. To compute C and Z, its indicator values will be compared to the values of BCI for the more corrupt countries. In other words, unlike the WGI index, the estimation does not rely on the idea that on average corruption has to be zero for this subsample. Instead, the estimated C will be positive to compensate for the relatively lower values of BCI that it is being compared to.

Standardization

Following the WGI index the expected value of corruption was standardized such that it has a mean of zero and a standard deviation of one. However, unlike the WGI index, this was done for the entire sample rather than on a yearly basis (3.A.4). Normalizing the yearly means would destroy the overall trend and invalidate comparisons over time (cf. section 3.5.4).

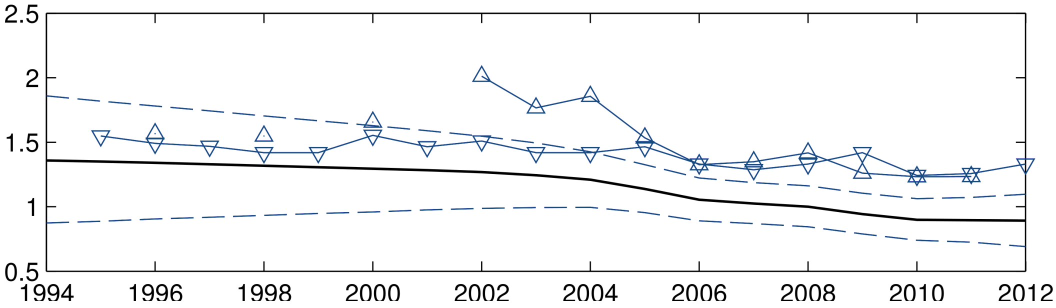
3.5 The Bayesian Corruption Indicator

Figure 3.3 plots of the Bayesian Corruption Indicator (BCI) for the USA, North Korea, South Africa and France alongside the values of the WGI and the CPI. Similar to the latter two, a *high* score for the BCI indicator means that there is *little* corruption.

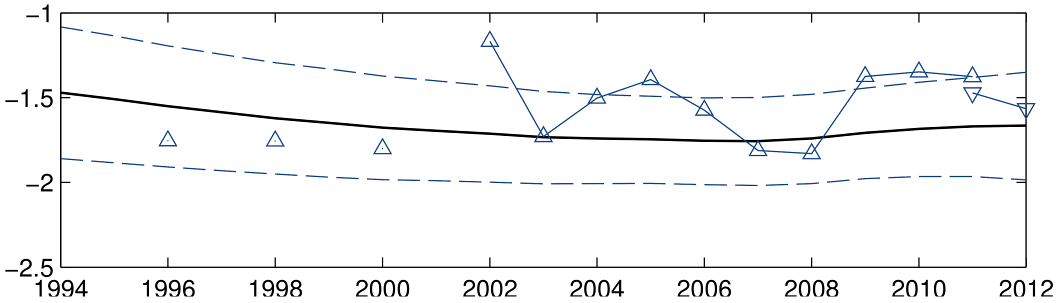
The estimation period is significantly expanded relative to the CPI and WGI. Even though the index was computed starting in 1984, we will only cover the results starting in 1995 seeing that there is only one source that goes back further. Even over this more limited period, the BCI index increases the size of the dataset by 40%. Also visible in figure 3.3 is the effect of an increase in the number of individual indicators on the size of the confidence intervals. For example, in 1995 there are 2 indicators that cover France and the uncertainty of the initial estimates is relatively large. As the number of indicators rises to 23 in 2010, this band narrows progressively. The subsequent drop to only five in 2012 once again increases the uncertainty.

Secondly, the BCI estimates are more stable (figure 3.3). Other corruption indicators (both individual and combined) have been criticized as being prone to small jumps in the data that have nothing to do with the level of corruption (Arndt and Oman, 2006; Treisman, 2007). By taking the past and future values of corruption into account, more information is used to discern between random measurement errors and actual changes in the level of corruption. As the time-dependency increases, more and more changes are filtered out and the resulting indicator becomes smoother (Kaufmann et al., 2004). On the other hand, as shown in the dynamic factor model of (Kaufmann et al., 2004), adding time-dependence is inadvisable when not corruption but the measurement errors are persistent. For this reason, section 3.7.5 adds persistent measurement errors to the state-space model, but finds that this model performs significantly worse than than the baseline model.

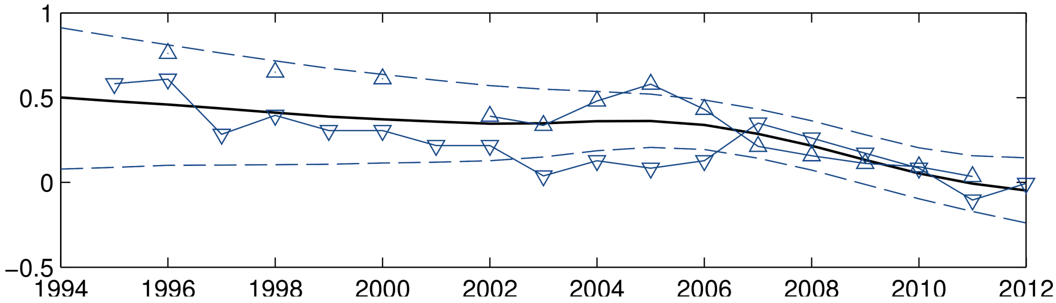
It should be reiterated that all available indicators are used in these estimations, meaning that the stability of the BCI is not the result of pre-selecting only the most



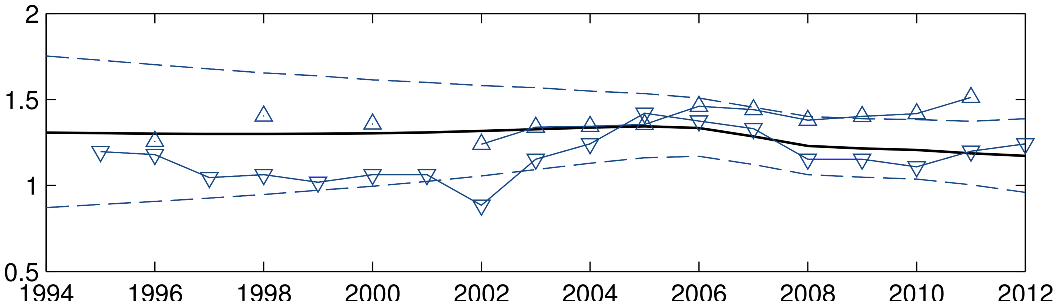
(a) United States of America



(b) Democratic Republic of Korea



(c) South Africa



(d) France

Figure 3.3: The BCI indicator over time
Plot of the BCI estimates, including 90% highest posterior density interval (dotted lines). Values of the standardized CPI (▽) and WGI (△) are also included.

stable individual indicators. In addition, the smoothness does not change when the indicators are grouped per source as in the WGI.

3.5.1 Correlations

Despite their methodological differences, the BCI, the WGI and the CPI give relatively similar predictions. Table 3.3 lists the pairwise correlations between the three indexes. Their overall correlation is very high. However, this is almost completely driven by their between-correlations (the correlation between the mean values for each country). The within correlation (between the demeaned values) on the other hand is significantly lower. In other words, while the choice of indicator might not have a large effect on the results in a cross-country study, this will change in time-series or panel studies (cf. section 3.5.4).

The reason why the within correlation is so low is related to the relative smoothness of the BCI estimate over time. As shown in figure 3.3 the small changes present in the WGI and CPI are identified as noise and filtered out of the BCI. The correlation of this noise over the different indicators is very low. Since the noise constitute a significant fraction of the total variation over time in WGI and CPI, the within correlation will also be low. On the other hand, the within correlation rises drastically when looking at large changes. This is shown in the last four columns where the correlation of the three indicators is computed over the significant changes as identified by either the rule of thumb or at the 10%, 5% or 1% significance level (cf. section 3.5.4).

Unfortunately, the pattern of missing values prevent us from comparing the goodness of fit of all variables simultaneously, as there are no observations where all indicators are available. Figure 3.4 shows the pairwise correlation of the BCI index with the individual indicators that went into the index. By far, most indicators are very highly positively or highly negatively correlated with the BCI index, meaning that a lot of the information contained in these indicators is used. Furthermore, the correlations are fairly similar over the different types of indicators, whether

Table 3.3: Pairwise correlations between BCI, WGI and CPI

	BCI - WGI	BCI - CPI	CPI - WGI
Total	0.948	0.956	0.970
Between$^{(a)}$	0.969	0.965	0.984
Within$^{(b)}$	0.352	0.202	0.347
Within - significant$^{(c)}$: RoT	0.803	0.946	0.892
Within - significant$^{(c)}$: 0.10	0.703	0.680	0.713
Within - significant$^{(c)}$: 0.05	0.760	0.815	0.758
Within - significant$^{(c)}$: 0.01	0.830	0.932	0.860

$^{(a)}$Between correlation is defined as the correlation between the means of each countries;
$^{(b)}$Within correlation is the correlation between the demeaned values of all countries.
$^{(c)}$Within correlation of significant changes over time as identified by the rule of thumb (RoT) or at the 10%, 5% or 1% significance level (cf. section 3.5.4).

they are surveys or expert opinions. In other words, unlike Arndt and Oman (2006) suggested, all types of sources are represented by the BCI index. Appendix 3.E provides a more detailed look by listing the R^2 of the measurement equation[5], ordered from highest to lowest, and decomposing it further into the within and between R^2. It reveals that the explained within variance is very low for most indicators, but this is mitigated by the fact that the between variation is the foremost source of variation in the BCI index. The between variance of the standardized index is 0.995, while the within variance is only 0.009. The individual indicator that does the best job capturing both within and between variation of the BCI is the Global Corruption Barometers' perception of corruption in parliament/legislature (GCB2).

3.5.2 Validity of the time-dependence parameter T_i

Finally, the values for parameter T give an indication of the necessity of the added time-dependence. As was explained earlier, setting $T = 0$ will reduce the BCI framework to that of the Worldwide Governance Indicators. However, the hypothesis that this parameter is zero is rejected for the vast majority of countries: 203 out of 219 at the 1% significance level. In fact, for most countries, T lies close to one (figure 3.5). Keep in mind that the values for T are restricted to lie within the [-1,1] interval,

[5]As each measurement equation is a univariate regression, the R^2 is equal to the square of the correlation.

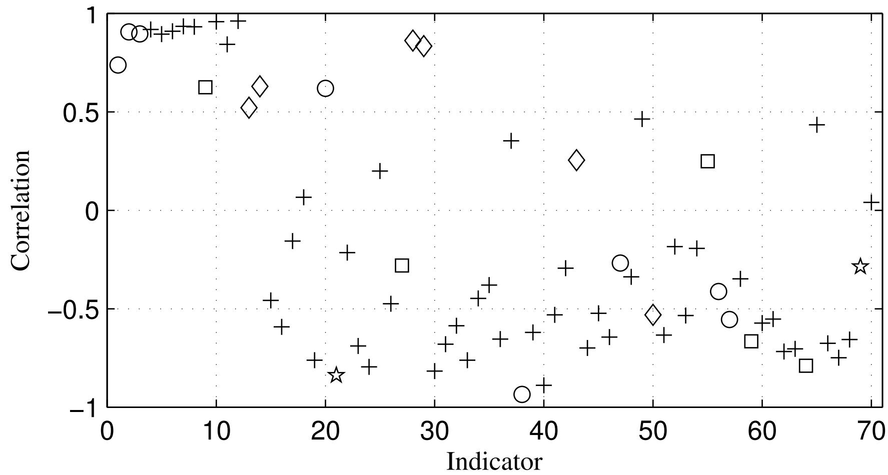


Figure 3.4: Correlations with individual indicators
Correlation of the 43 individual indicators of corruption with the BCI. ○ mark indicators from commercial sources; + from firm-level surveys; ◇ from governmenental and multilateral sources; □ from household surveys; and ☆ from NGOs.

to ensure that corruption is a stable, non-explosive time series and the steady-state variance of corruption is non-negative.

The formal test of the validity of T_i involves comparing each models' marginal probability. Similar to the likelihood ratio, marginal probability of a model expresses how likely the data is, given that it was generated using this model. The ratio between the two marginal probabilities, the bayes factor, will be greater than one if the model in the numerator better suits the data (Carlin and Louis, 2000). The computation of the marginal probability of each model is described in 3.A.6. The bayes factor confirms that the time-dependence is a valid addition to the model: according to Kass and Raftery (1995), a value of more than 5 is decisive evidence in favor of the baseline model.

$$\ln[BF_{baseline,T=0}] = \ln\left[\frac{p(y|baseline)}{p(y|T=0)}\right] = 443$$

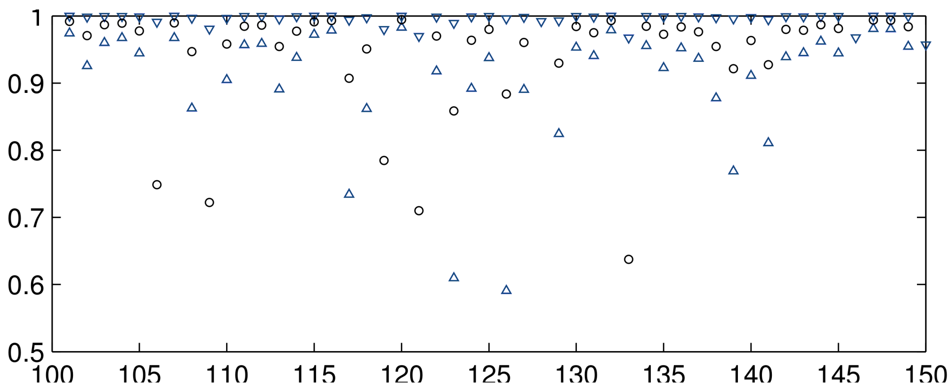


Figure 3.5: Plot of mean values of T and their 95% highest posterior density interval
Plot of expected values parameter T_{100} to T_{150} (circles) and their 95% highest posterior density intervals (triangles).

3.5.3 Reliability

By using the time-structure present in the data, the estimates of corruption can be made with greater precision and certainty. Table 3.4 lists the average size of the confidence intervals and highest posterior density intervals of the three combined indicators over different samples and using different standardization methods.

In order to make the right comparison between the average standard deviations of the three indicators, it is important to compare them over the same sample and using the same standardization techniques. The BCI indicator goes farther back in time than the WGI, covering countries and time periods for which there is less certainty. Once corrected for this, it becomes clear that the average standard deviation of the BCI is smaller than that of the WGI. Using a yearly normalization for the BCI index does not change this conclusion. The WGI's confidence interval is smaller than that of BCI in less than 10% of the sample.

The correlation coefficient between the two standard deviations is 0.607, meaning that for the most part the standard deviation of BCI and WGI follow the same pattern. The differences are caused the adding the time parameter T, the individual indicators used (even though they come from the same sources), and the fact that α and ε are no longer assumed to be multivariate normal distributed (eqn. 3.5).

The CPI on the other hand scores worse than both the BCI and WGI, regardless of normalization and sample size. Moreover, its confidence bounds follow a completely different pattern: the correlation coefficient with BCI and WGI is -0.031 and 0.243, respectively.

Table 3.4: Average standard deviation of BCI, WGI and CPI

Standardization$^{(A)}$	Sample	BCI	WGI	CPI
None	Total	3.226	0.236	1.087
Total	Total	0.249	0.236	0.484
Total	WGI	0.156	0.236	0.258
Total	CPI	0.163	0.182	0.484
Yearly	Total	0.247	0.236	0.531
Yearly	WGI	0.157	0.236	0.262
Yearly	CPI	0.163	0.182	0.531

Average standard deviation of the BCI, WGI and CPI. $^{(A)}$*None:* indicator values are used without normalization; *Total:* all mean values are normalized to mean of zero and standard deviation of one (cf. BCI); *Yearly:* yearly mean values are normalized to mean of zero and standard deviation of one (cf. WGI).

3.5.4 Significant changes in corruption

Table 3.5: Changes in the level of corruption between 2000 to 2010

Significance	Deteriorated (BCI decreased)		Improved (BCI increased)
1%	Italy$^{RoT,\,WGI}$	Greece$^{RoT,\,WGI}$	Georgia$^{RoT,\,WGI}$
	Czech Rep.^{RoT}	Hungary^{RoT}	Saudi Arabia^{RoT}
5%	Kuwait	South Africa	Gambia
	Colombia	USA	Iraq
	Iceland	Venezuela^{WGI}	Macedonia^{WGI}
	Slovakia		Qatar^{WGI}
10%	Algeria	Maldives	Bahrain
	Argentina	Moldova	Indonesia
	Austria	Russia	Liberia^{WGI}
	Brazil	Slovenia	Oman
	Bulgaria	Spain	Palestine
	Croatia	Ukraine	Zimbabwe
	Madagascar		
not	Eritrea^{WGI}		Rwanda^{WGI}
	Great Britain^{WGI}		Serbia^{WGI}
			UAE ^{WGI}

List of countries whose level of corruption changed significantly between 2000 and 2010. ^{RoT} indicates whether the change was detected using the rule of thumb and ^{WGI} using WGI's data.

Before comparing any countries, the question should be asked whether the underlying indicators allow comparisons over time. For example, the use of ranked data in the old methodology of the CPI limited its use to cross-sectional studies. However, an analysis of the survey questions listed in appendix 3.B reveals that this is not the case for the BCI index. For example, the frequency of payments/gifts when dealing with the taxes and tax collection (BPS14), the perception of corruption in the judiciary (GCB14) and the fraction of people that identify corruption as the most serious problem facing the country (VAB1) are all variables that can be meaningfully compared over time.

In order to see whether the level of corruption has changed over time, Kaufmann et al. (2004) suggest a rule of thumb based on dynamic factor model that takes into account persistence in corruption as well as in the measurement errors. The rule of thumb is that if the 90% confidence intervals overlap, the change is deemed big enough to be significant. The problem is that this ignores the time structure in the corruption data. If corruption did not depend on its previous values, this approximation would return relatively good results. However, because of the high level of persistence this rule ends up making a lot of type I errors by labeling significant changes as not significant.

Using the data from the Gibbs sampler, it is possible to test whether the change in corruption is significant. If in more than 95% of the drawn values of α a country's level of corruption decreases, this change is significant at 5% significance level. This can be extremely useful given the increased importance of changes in governance in for example the allocation of international aid (Kaufmann et al., 2004; Arndt and Oman, 2006).

Table 3.5 lists those countries for which there was a significant change in the level of corruption between 2000 and 2010 at 1, 5 and 10% significance levels. It also indicates which of these changes would have been identified using the rule of thumb and using the WGI.

Even though the rule of thumb uses the 90% confidence intervals, it is only able to identify the changes in the level of corruption at the 1% level. However, it does not necessarily capture all of the changes at the 1% level. For example, it misses almost 40% between 2005 and 2012 (9 out of 23).

The results using the WGI on the other hand differ strongly from those of the BCI. Using WGI data, some of the changes from 1,5 and 10% are identified (7 out of 36), but 5 shifts are also spuriously identified. This demonstrates the fact that despite their high overall correlation, BCI and WGI will come to significantly different conclusions when used for making comparisons over time.

Worldwide trend

The evolution of the average worldwide level of corruption can be investigated in a similar way. As figure 3.6 makes clear, a simple rule of thumb would not be able to find any significant evolution in the average level of corruption. However, using the results from the Gibbs sampler, the overall decrease between 1995 and 2012 is significant at 5% and the drop between 2005 and 2012 even at 1%. The small increase between 1995 and 2005 on the other hand is not significant.

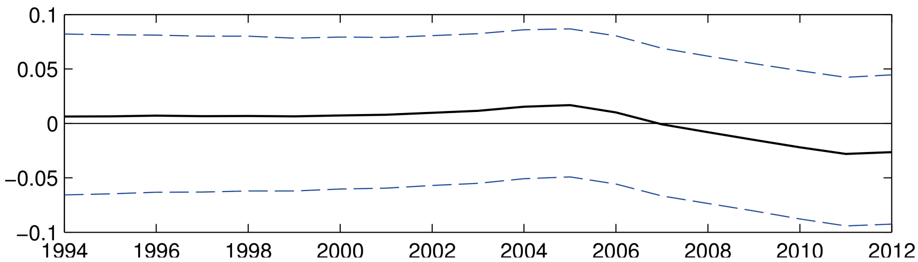


Figure 3.6: Worldwide trend in corruption values
Plot of the yearly mean value of the BCI index and its 95% highest posterior density interval (dotted lines).

Rankings

Analogous to the comparisons over time, the index also allows us to identify significant differences between countries. Country A has significantly less corruption (a higher score) than country B at the 1% level if its score is higher in more than 99% of the drawn values. For example, using the cross-correlated indicator 16,841 significant differences can be found between 210 countries/regions in 2010 at the

Table 3.6: The 15 best and worst ranked countries in 2012

rank	country	BCI	rank	country	BCI
1	Denmark	2.26	27	Myanmar	-1.64
1	New Zealand	2.21	27	Somalia	-1.62
1	Finland	2.12	27	Korea, P.D.R.	-1.59
1	Sweden	2.09	27	Iraq	-1.42
2	Singapore	2.10	27	Afghanistan	-1.41
2	Netherlands	1.97	27	Angola	-1.37
2	Switzerland	1.94	27	Haiti	-1.34
2	Australia	1.93	27	New Caledonia	-0.97
2	Iceland	1.87	26	Chad	-1.40
2	Norway	1.87	26	Turkmenistan	-1.37
2	Canada	1.87	26	Sudan	-1.34
3	United Kingdom	1.57	26	Equatorial Guinea	-1.30
3	Austria	1.53	26	Congo, D.R.	-1.28
4	Luxembourg	1.86	26	Burundi	-1.27
4	Hong Kong	1.75	26	Zimbabwe	-1.21

This table lists the 15 countries with the lowest (left) and highest (right) level of corruption. The ranking is based solely on significant differences in the cross-correlated BCI scores at the 5% significance level.

1% level. Using the WGI and the rule of thumb only 15,298 can be identified: 2,085 are no longer significant while 542 are significant using WGI but not using BCI.

Table 3.6 uses these differences to rank countries according to the following rule: country A is ranked higher than country B if, and only if, its level of corruption is significantly lower than that of B or than that of a country whose level of corruption is not significantly different from that of B. In other words, a country has rank x if it is significantly less corrupt than at least one country with rank $x+1$.

The main advantage of this ranking relative to one based only on the values of the index is that it ignores the insignificant differences between countries. As Høyland et al. (2012, p.2) pointed out, simply using the scores of an indicator to rank countries can falsely give the impression of a clear-cut ordering of countries. Using a ranking based on significant differences removes some of the randomness of rankings, making them a more useful policy analysis tool.

3.6 Selection bias

As a number of authors, including Kaufmann et al. (2010), have made clear, the initial values of the WGI and the CPI potentially suffer from selection bias issues. Because early commercial corruption indicators focused on countries their clients are interested to trade in, they were more likely to cover countries with high levels of GDP and/or low levels of corruption. Consequently, the selection of countries covered by CPI and WGI could depend on the level of corruption and GDP. However, because the BCI index uses the time structure in the data, it is able to provide an estimate of the level of corruption for all countries starting in 1984, avoiding this selection bias problem.

Additionally, it is possible to use the BCI values to formally test the existence of a selection bias in the WGI and the CPI. Defining D^{WGI} and D^{CPI} as dummy variables indicating whether or not a country is covered by WGI or CPI (1 if it is, 0 otherwise), equation 3.9 is tested. Following Treisman (2007), real level of GDP from the Penn World tables was also included in equation 3.10 (Heston et al., 2012).

$$\bar{D}^A_{i,t} = c + \beta_{BCI} * BCI_{i,t} + \psi_{i,t} \tag{3.9}$$

$$\bar{D}^A_{i,t} = c + \beta_{BCI} * BCI_{i,t} + \beta_{GDP} * GDP_{i,t} + \psi_{i,t} \tag{3.10}$$

$$D^A_{i,t} = \begin{cases} 1 & \text{if } \bar{D}^A_{i,t} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\psi_{i,t} \sim N(0,\sigma)$$

with $A \in \{CPI, WGI\}$ and $t \in [1995, 2011]$.

For each year, these equations were estimated using a Bayesian Gibbs sampler algorithm. The uncertainty of the corruption estimate was taken into account by using a different generated value of the BCI index in each iteration of the Gibbs sampler. Figure 3.7 summarizes the results by plotting the coefficients on BCI and GDP for each year (the complete tables are in 3.C). These show that the availability of the

WGI is influenced by the level of corruption in 1996, but only at the 10% significance level. For all other years, only the level of GDP plays a role. However, the GDP data is only available for 189 countries and almost perfectly matches the countries covered by WGI, meaning that these results should be interpreted with the necessary caution.

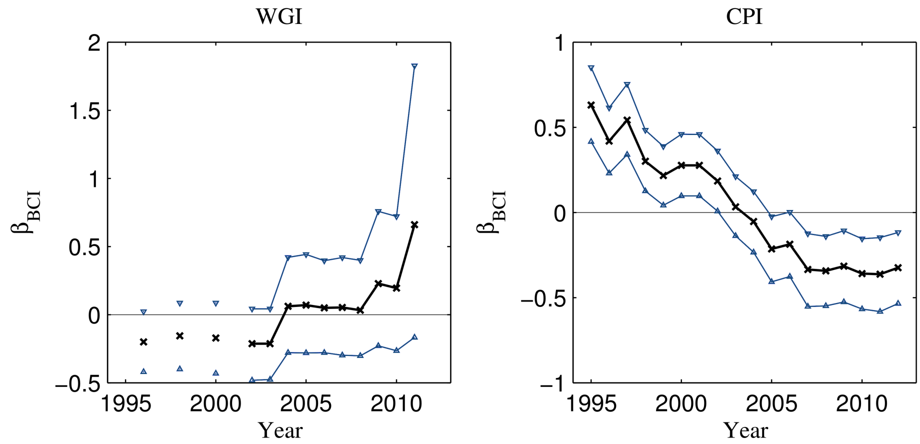
The pattern is reversed in the case of the CPI where, with the exception of 2004, there is evidence that its selection is significantly influenced by the level of corruption in all years. Initially CPI tends to cover less corrupt countries more, but from the mid-00's this pattern is turned upside down. In addition, the higher the level of GDP, the more likely the country will be covered by the CPI. Regardless of the direction of the effect, the fact that the level of corruption influences whether a country is covered or not strongly cautions against using the (early values of the) CPI in statistical research as they are likely to produce biased results.

3.7 Robustness checks

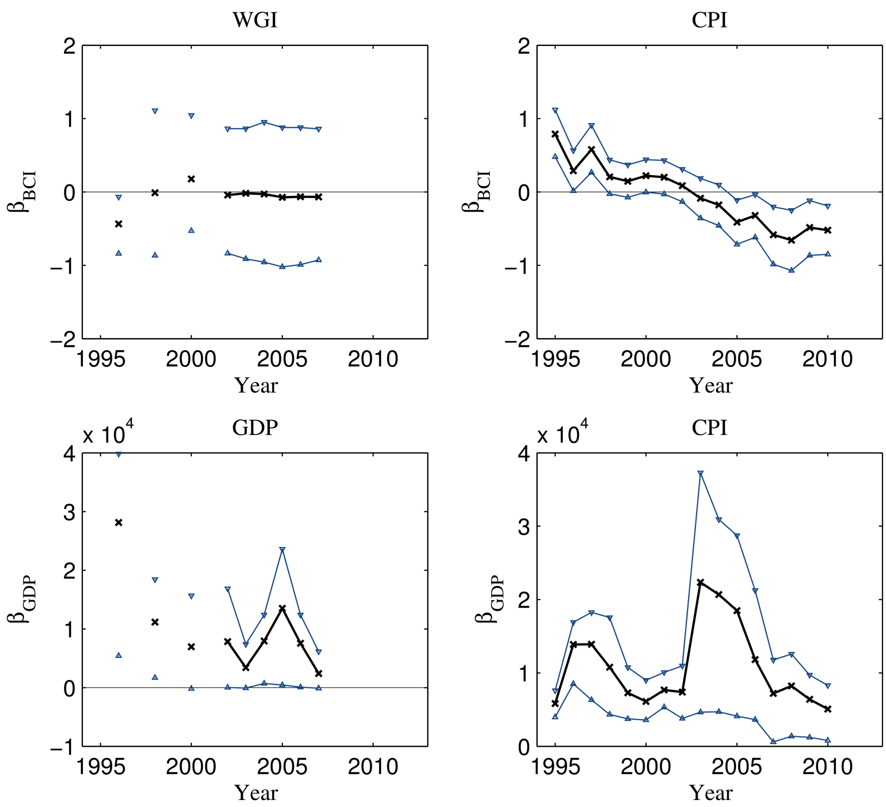
3.7.1 Stability of the parameters

In contrast to the Worldwide Governance Indicators, the parameters of the measurement equation are assumed to be constant over time (eqn. 3.5). In order to test the validity of this assumption, the model was run for five (overlapping) periods to see whether the parameters remain stable: 1995-2005, 1997-2007, 2000-2010 and 2002-2012. Because many of the indicators are only available for a couple of years, the dataset was restricted to the five variables that have sufficient observations in all periods to run the model: EIU, PRC, PRS, WCY and WMO.

Figure 3.8 shows the empirical distribution function of the slope parameter Z for the different sub-periods. It clearly shows that while there are small differences, these are not significant for any of the variables. This is confirmed in table 3.7 where the results of a formal, two-sided test are presented. The null-hypothesis that the parameters of the subset are equal to those of the baseline model can never be



(a) Corruption only



(b) Including GDP

Figure 3.7: Selection bias in CPI and WGI
Plot of the expected value (black crosses) and 95% highest posterior density interval (blue triangles) of the coefficient of BCI and GDP in equations 3.9 and 3.10 for different years.

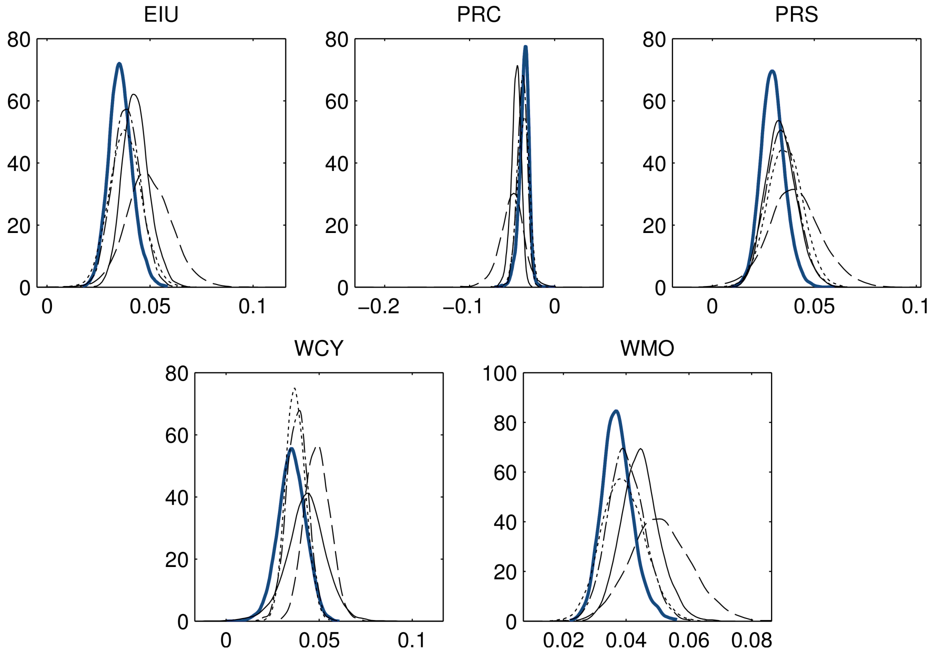


Figure 3.8: Stability slope parameter Z
Plot of the parameter values of Z estimated over different time periods: 1995-2012 (bold blue line), 1995-2005 (full line), 1997-2007 (dashed line), 2000-2010 (dash-dotted line), 2002-2012 (dotted line).

rejected.[6] In other words, the assumption that these coefficients are constant over time should not affect the resulting indicator.

3.7.2 Keeping T_i fixed for all countries

In contrast to the parameters of the measurement equation,(C Z and H) the time-dependence parameter (T_i) can be different for each country. Considering the cost in degrees of freedom this imposes, this section considers setting T_i equal to τ for each country. Re-estimating the model under this restriction returns very similar expected values (table 3.8). The within correlation further increases to 98% when only the values from 1995 onwards are compared (removing the period when only a couple of indicators are available). The effect of keeping T fixed on the standard deviations on the other hand is stronger: their average size increases with 30% while their overall correlation with those of the baseline model is only 0.782.

[6]If the $H_0 : \beta_s = \beta_o$, $H_1 : \beta_s \neq \beta_o$ and $E[\beta_s] < E[\beta_o]$, the p-value is computed as: $2 * p(\beta_s > \beta_o)$.

Table 3.7: Stability of the slope parameter Z

		Overall	1995	1997	2000	2002
PRS	Mean	0.03	0.034	0.041	0.034	0.036
	St.dev.	−0.006	−0.008	−0.013	−0.008	−0.009
	p-value[a]		0.683	0.435	0.665	0.565
WMO	Mean	0.037	0.045	0.051	0.04	0.039
	St.dev.	−0.005	−0.006	−0.01	−0.006	−0.007
	p-value[a]		0.344	0.199	0.711	0.829
PRC	Mean	−0.036	−0.045	−0.05	−0.038	−0.037
	St.dev.	−0.006	−0.006	−0.015	−0.006	−0.007
	p-value[a]		0.243	0.323	0.756	0.905
EIU	Mean	0.036	0.043	0.049	0.038	0.038
	St.dev.	−0.006	−0.006	−0.011	−0.007	−0.008
	p-value[a]		0.383	0.258	0.75	0.804
WCY	Mean	0.035	0.043	0.049	0.038	0.037
	St.dev.	−0.007	−0.011	−0.007	−0.006	−0.005
	p-value[a]		0.529	0.166	0.755	0.824

[a]p-value of the 2-sided test of the equality of the overall parameter to the parameter of the 1995/1997/2000/2002 subset.

When the restriction is split up into $T_i = T_j$, $\forall i \neq j$ almost 10% is rejected (at 5% significance level) using the values from the baseline model. Moreover, comparing each value of T_i from the baseline model with the posterior density of τ, more than a quarter is found to be significantly different at 5% significance level. In other words, similar to what was found in table 3.2, the persistence in the level of corruption is high for the majority of countries, but there are some where the level of corruption changes more quickly. While the Bayes factor is lower for models with more variables (all other things equal), it nevertheless finds that the baseline model outperforms the model where $T_i = \tau$:

$$\ln[BF_{baseline,T_i=\tau}] = 1566.$$

In addition to keeping T_i fixed, one could also impose a unit root on the BCI index and estimate the model in first differences. There are however a number of problems with the corruption data that undermine the usefulness of this approach. The most important one is that the variation in the corruption data is mostly cross-sectional, which is lost when using first differences. In addition, because not all indicators are available on a yearly basis, 29 of the 69 indicators drop from the dataset and

Table 3.8: Correlations between the different corruption indicators

			Overall	Within	Between
BCI	-	$\text{BCI}_{T_i=\tau}$	0.990	0.683	0.993
BCI	-	$\text{BCI}_{Grouped}$	0.968	0.713	0.971
WGI	-	$\text{BCI}_{Grouped}$	0.975	0.455	0.977
BCI	-	$\text{BCI}_{Block-diag}$	0.974	0.753	0.977
$\text{BCI}_{Grouped}$	-	$\text{BCI}_{Block-diag}$	0.998	0.964	0.999
BCI	-	$\text{BCI}_{persistence}$	0.970	0.661	0.963

Between correlation is defined as the correlation between the average values of countries; The within correlation is the correlation between the demeaned values of all countries.

the number of non-missing observations is reduced by 30%. As a result, the within correlation of the first-differences model with WGI, CPI and the baseline BCI is only 0.12, 0.05 and 0.25, respectively.

3.7.3 Grouping indicators per source

The BCI index is computed under the assumption that the indicators do not have a shared measurement error: $E[\varepsilon^{(k)}, \varepsilon^{(m)}] = 0,\ \forall\ k \neq m$. However, it is unlikely that indicators coming from the same source do not share a certain ideological or methodological bias. While ignoring the cross-correlation will not skew the drawn values of Z, C and H in step A (figure 3.2)[7], the same cannot be said of the estimation of α in step B.

To control for cross-correlation, the WGI first groups all observations coming from the same source using a simple average. While grouping the indicators avoids the cross-correlation problem, it introduces two potential new ones. Firstly, using the average ignores the underlying variability of the individual indicators (a generated variable bias). For example, when one sub-indicator gives a score of 0/10 and another 10/10, the average is seen as equally reliable as a source where both sub-indicators rate 5/10. As a result, the size of the confidence intervals would be driven down.

In addition, the composition of the indicators per source changes from year to year

[7]The measurement equation is a Seemingly Unrelated Regression model with identical independent variables, meaning that its estimates are mathematically identical to running k independent linear regressions.

because of missing values, especially in firm and household surveys. This means that a change in the average value of the indicator could simply be the result of a different composition per source. This measurement error would in turn increase the size of the confidence intervals.

To illustrate, the index was recomputed using the average values per source. Table 3.8 shows that the grouped indicator lies somewhere in between the BCI and WGI values. The main difference when grouping data lies in the size of their predicted standard errors. They are on average 15% smaller for the grouped indicator: 0.210 (grouped) versus 0.245 (BCI). This indicates that the negative effect of the generated variable bias on size of the confidence intervals outweighs the positive effect of the measurement error and loss of information.

3.7.4 *H* block-diagonal

A more straightforward solution is to simply allow variables from the same source to have a shared measurement error: $E[\varepsilon^{(k)}, \varepsilon^{(m)}] \in \mathbb{R}$, if k and m originate from the same source. The effect on the estimation procedure remains limited. When drawing values for α the variance matrix H is now block-diagonal rather than diagonal (step A, figure 3.2). In addition, the diagonal elements of H can no longer be drawn separately for each indicator (step B). Instead, they are now drawn per cluster in the style of a seemingly unrelated regression model[8].

The many missing variables do complicate matters, in that the clusters cannot simply be defined as indicators coming from the same source. Some indicators do not share enough observations to enable them to be grouped. More specifically, there are 17 such variables, six of which are treated as individual data and others that can still be clustered in smaller groups. The complete list can be found in the last column of 3.B where the exceptions are marked in bold.

Table 3.8 lists the correlations between the block-diagonal, the original and the grouped index and the WGI. It shows that the block-diagonal indicator is close to

[8]This approach differs from that of Høyland et al. (2012), who account for group-specific shocks by using a group-specific error term in addition to an idiosyncratic one.

perfectly correlated with the grouped indicator on the overall as well as the between and the within level. However, as with the grouped indicator the real difference lies in the predicted standard errors. The average standard deviation of the block-diagonal indicator is 0.2409, which is larger than that of the grouped indicator but smaller than the baseline model.

To put the changes in significance in perspective, 3.D lists the significant changes in corruption between 2000 and 2010 according to the block-diagonal indicator. Only 3 of the 47 countries appear in the same place as in table 3.5. Interestingly, quite a few countries whose change was identified using the WGI index but not using the BCI index are now also found to have changed significantly. In other words, ignoring the cross-correlation in the corruption indicators does have an impact on the combined indicator.

Finally, the majority of the estimated cross-correlations is significant at the 5% level (155 of the 195). The validity of the added cross-correlation is confirmed by the bayes factor: $\ln[BF_{baseline,Hdiag}] = -2139$.

3.7.5 Persistent measurement errors

An alternative adjustment to the measurement equation is to allow the individual indicators to make persistent measurement errors, in other words letting some indicators persistently under- or overestimate the level of corruption. To that end, the equation 3.5 is adjusted to include an error term following an AR(1) process:

$$y_{i,t} = C + Z\alpha_{i,t} + \varepsilon_{i,t} \tag{3.11}$$

$$\varepsilon_{i,t} = D\varepsilon_{i,t-1} + \xi_{i,t} \tag{3.12}$$

$$\xi_{i,t} \sim N(0,H) \tag{3.13}$$

with D a diagonal matrix and $D_{(\kappa,\kappa)} = 0$ if the indicator y^{κ} is not available on a yearly basis. The changes to the estimation procedure are outlined in 3.A.3.

Similar to what happened when $T_i = \tau$, the expected values of the BCI change little when adding persistent measurement errors (table 3.8). The standard deviations on the other hand are more affected: they increase by 30% over the entire sample, and even 60% when comparing from 1995 onwards.

While the parameter values of D seem to indicate that the persistence measurement errors are a valid addition to the model (Figure 3.9), the Bayes factor nevertheless prefers both the baseline and the cross-correlated model to the model with persistent measurement errors.

$$\begin{aligned}\ln[BF_{baseline,persistence}] &= 4865 \\ \ln[BF_{Hdiag,persistence}] &= 7079\end{aligned}$$

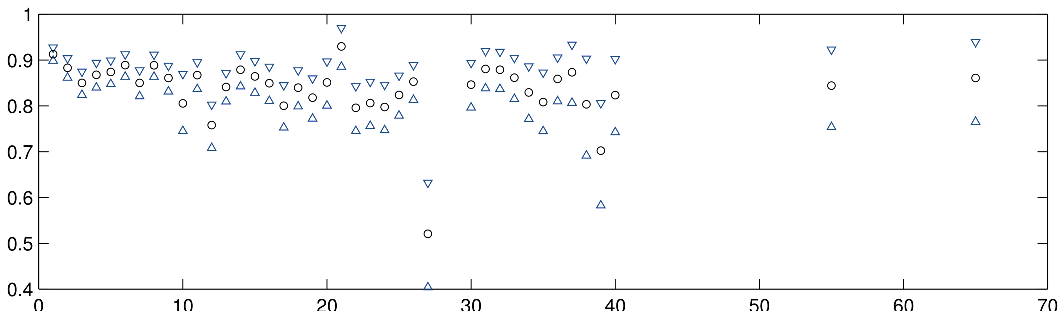


Figure 3.9: Persistence of the measurement errors ($D_{\kappa,\kappa}$)
Plot of the expected value for the diagonal elements of D (circles) and their 95% highest posterior density interval (blue triangles).

3.8 Conclusion

The Bayesian Corruption Indicator (available at `www.sherppa.be`) improves on the existing corruption indicators in a number of ways.

From a practical point of view, the BCI indicator can predict the level of corruption with greater certainty and increases coverage. The possibility of capturing the shared measurement error of certain corruption indicators further enhances the reliability of the estimates and is found to be the model that fits the corruption data best. Allowing some indicators to persistently under- or overestimate the actual level of corruption has little influence on the expected values of the BCI index, although the effect on the standard deviations is more substantial.

By taking the time-aspect into account, the model more effectively filters out random measurement errors, leading to more stable estimates. Most importantly, the BCI index does not suffer from the serious selection bias issues that plague the Corruption Perceptions Index. Finally, because the estimation of the BCI returns the entire distribution of corruption, it makes it possible to say whether or not the level of corruption significantly increased or decreased over time with greater accuracy.

From an estimation point of view, the underlying assumptions of the BCI model are explicitly stated, making it a very transparent approach. The combination with the solution to missing data points also eliminates the need for additional assumptions, imputations or sub-level aggregations, further increasing the objectivity of the index.

Lastly, from a theoretical perspective, the parameter values and Bayes factors clearly indicate that the time-dependence of corruption and cross-correlated errors are valid additions to the state-space model. Moreover, in most countries the level of persistence is very close to one, meaning that the utmost caution has to be used when regressing corruption on other non-stationary series. It also invalidates any regressions on stationary data as they should lead to insignificant results.

References

Arndt, C. and Oman, C. (2006) Uses and abuses of governance indicators. OECD Publishing.

Carlin, B.P. and Louis, T.A. (2000) *Bayes and empirical Bayes methods for data analysis.* Chapman and Hall/CRC, Boca Raton.

Carter, C.K. and Kohn, R. (1994) On gibbs sampling for state space models. Biometrika 81(3):541–553.

Durbin, J. and Koopman, S. (2012) *Time series analysis by state space methods, 2nd edition.* Oxford University Press, Oxford.

Givens, D. (2013) Defining governance matters: A factor analytic assessment of governance institutions. *Journal of Comparative Economics* 41:1026–1053.

Heston, A., Summers, R. and Aten, B. (2012) Penn world table version 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Høyland, B., Moene, K. and Willumsen, F. (2012) The tyranny of international index rankings. *Journal of Development Economics* 97(1):1–14.

Kaufmann, D., Kraay, A. and Zoido-Lobatón, P. (2004) Governance matters III: Governance indicators for 1996, 1998, 2000 and 2002. *World Bank Economic Review* 18:253–287.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2005) Governance matters IV: Governance indicators for 1996-2004. The World Bank Technical Report WPS3630.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2007a) The worldwide governance indicators: Answering the critics. World Bank Policy Research Paper 4194.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2007b) Governance matter VI: Aggregate and individual governance indicators for 1996-2006. World Bank Policy Research Paper 4280.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010) The worldwide governance indicators - methodology and analytical issues. The World Bank.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *Journal of American Statistical Association* 90(430):773–1795.

Kim, C.J. and Nelson, C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* MIT Press, Cambridge.

Koop, G., Poirier, D.J. and Tobias, J.L. (2007) *Bayesian econometric methods.* Cambridge University Press, New York.

Lambsdorff, J.G. (2005) The methodology of the 2005 corruption perceptions index. University of Passau Technical report.

Lancaster, T. (2004) *Introduction to modern Bayesian econometrics.* Blackwell, Oxford.

Roca, T. (2011) Measuring corruption: perception surveys or victimization surveys. Groupe d'économie du développement Lare-Efi Working paper DT/167/2011.

Saisana, M. and Saltelli, A. (2012) Corruption Perceptions Index 2012: Statistical Assessment. European Commision JRC Scientific and Policy Reports.

Transparency International. CPI index - in detail.

Treisman, D. (2007) What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science* 10:211–244.

Verbeek, M. (2010) *A guide to modern econometrics* John Wiley & Sons, Chichester.

Appendices

3.A Estimation

3.A.1 Priors

Flat probabilities

Because the state-space model is estimated in a Bayesian framework, it is necessary to specify the prior distribution of the parameters. However, since there is no prior information available on the parameters of the measurement equation, Z, C and $log(H)$, flat probabilities are used for these variables. This means that all values in $\mathbb{R}$ are equally probable. It is important to note that the WGI or CPI indexes cannot be used as sources of prior information, seeing that they are based on the same data sources used in the estimations.

$$p(Z) \propto 1_{(k,1)} \tag{3.14}$$

$$p(C) \propto 1_{(k,1)} \tag{3.15}$$

$$p(\log(H)) \propto \mathbb{I}_k \tag{3.16}$$

with $1_{(k,1)}$ a $(k \times 1)$ vector of ones, $\mathbb{I}_k$ an identity matrix of size $(k \times k)$ and k the total number of individual indicators.

For the state equation, there is a prior restriction on T_i that its absolute value does not exceed one, for all countries i. This ensures that $\alpha_{i,t}$ is a non-explosive time series, without precluding non-stationary series.

$$p(T) = 0.5 * \mathbb{1}_{|T| \leq 1} \tag{3.17}$$

Finally, as an identifying assumption, the variance of the state equation, Q, is set to one.

Actual probabilities

As a robustness check, the model was also estimated using actual, but uninformative prior probabilities. However, the resulting indicator did differ from the one using flat probabilities.

$$Z|H \sim N(0, \frac{1}{2}); \tag{3.18}$$

$$C|H \sim N(0, \frac{1}{2}); \tag{3.19}$$

$$H \sim iWish(8,4); \tag{3.20}$$

$$T|Q \sim N(0, \frac{1}{2}Q). \tag{3.21}$$

with $iWish$ the inverse Wishart distribution.

To see why these probabilities are uninformative, consider first of all that the yearly change in corruption has been set to have a standard deviation of one ($Q = 1$). However, we expect the level of corruption to have a high persistence, which implies that the standard deviation of worldwide α values will be greater than one. Secondly all individual indicators are normalized to mean 0 and standard deviation one. Substituting this into the measurement equation (3.5), it follows that the parameter values C, Z should be smaller than one if the indicator is to be informative. Similarly the measurement error should have a standard deviation smaller than one.

3.A.2 Gibbs sampler

As was explained, the Gibbs sampler allows us to split the estimation process up in two main blocks, which can then be further divided into a number of easily solvable subroutines (Kim and Nelson, 1999):

A. Conditioning on the values for $\alpha = (\alpha_{1,0}, \ldots, \alpha_{1,n}, \ldots, \alpha_{p,1}, \ldots, \alpha_{p,n})'$, the state and measurement equation are reduced to simple linear regressions:

$$
\begin{aligned}
p(T|\alpha,Q) &\propto .5 * \mathbb{1}_{|T|\leq 1} * N[b_{OLS}(\alpha_t, \alpha_{t-1});\ (\alpha'_{t-1}\alpha_{t-1})^{-1}Q] && (3.22)\\
p(Z,C|\alpha,y,H) &\propto N(b_{OLS}(y,[\alpha,\mathbf{1}]),([\alpha,\mathbf{1}]'[\alpha,\mathbf{1}])^{-1}H] && (3.23)\\
p(H|\alpha,y) &\propto iWish[e'e;\ n] && (3.24)
\end{aligned}
$$

with $e \equiv y - T\alpha$ and $b_{OLS}(Y,X) \equiv (X'X)^{-1}(X'Y)$

B. Conditional on the parameters of the state and measurement equations, the probability of the 'true' level of corruption can be computed and drawn from using the Carter and Kohn (1994) simulation smoother.

- *The Kalman filter*

  Starting from a wild guess, $p(\alpha_0) = N(0,\infty)$, the following equations are iteratively solved for $t = 1$ to $t = n$:

$$
\begin{aligned}
a_{t|t} &= E(\alpha_t|y_1,\ldots,y_t)\\
&= T * a_{t-1|t-1} + \kappa(y_t - C - ZTa_{t-1|t-1}) && (3.25)\\
p_{t|t} &= V(\alpha_t|y_1,\ldots,y_t)\\
&= p_{t|t-1} + \kappa Z p_{t-1|t-1} && (3.26)
\end{aligned}
$$

  with $\kappa = p_{t|t-1}Z'(Zp_{t|t-1}Z' + H)^{-1}$; and $p_{t|t-1} = Tp_{t-1|t-1}T' + Q$.

- *Simulation smoother*

  The simulation smoother algorithm is used to draw values for α for each country one at a time. Starting from the last iteration of the Kalman filter, draw $\hat{\alpha}_n$ from $N(a_{n|n};\ p_{n|n})$ and iterate backwards from $t = n-1$ to $t = 1$:

$$\begin{aligned} a_{t|n} &= E(\alpha_t|y_1,\ldots y_n) \\ &= a_{t|t} + \varsigma(\hat{a}_{t+1|n} - Ta_{t|t}) \qquad (3.27) \\ p_{t|n} &= V(\alpha_t|y_1,\ldots y_n) \\ &= p_{t|t} + \varsigma(p_{t+1|n} - Tp_{t|t}T' - Q)\varsigma' \qquad (3.28) \end{aligned}$$

with $\varsigma = p_{t|t}T'p^{-1}_{t+1|t}$; and $\hat{a}_{t+1|n}$ a random draw from $N(a_{t+1|n};\ p_{t+1|n})$.

3.A.3 Persistent measurement errors

In order to estimate the model including persistent measurement errors, the model first has to be rewritten in the standard state-space form:

$$y_{i,t} = C + [Z\,\mathbb{I}_k]\begin{bmatrix} \alpha_{i,t} \\ \varepsilon_{i,t} \end{bmatrix} \qquad (3.29)$$

$$\begin{bmatrix} \alpha_{i,t} \\ \varepsilon_{i,t} \end{bmatrix} = \begin{bmatrix} T_i & 0_{(1,k)}) \\ 0_{(k,1)} & D \end{bmatrix}\begin{bmatrix} \alpha_{i,t-1} \\ \varepsilon_{i,t-1} \end{bmatrix} + \begin{bmatrix} \mu_{i,t} \\ \xi_{i,t} \end{bmatrix} \qquad (3.30)$$

It can subsequently be estimated much in the same way as described above with equation 3.29 and 3.30 as the measurement and state equation, respectively. One important difference is that the parameter values Z and C are now computed and drawn conditionally on α, H and D using the following equation:

$$\begin{aligned} y_{i,t} - Dy_{i,t-1} &= (1-D)\,C - Z(1_{(k,1)}\,\alpha_{i,t} - D\,1_{(k,1)}\,\alpha_{i,t}) + \varepsilon_{i,t} - D\,\varepsilon_{i,t-1} \\ &= (1-D)\,C - Z(1_{(k,1)}\,\alpha_{i,t} - D\,1_{(k,1)}\,\alpha_{i,t}) + \xi_{i,t} \end{aligned}$$

with $\xi_{i,t} \sim N(0,H)$.

3.A.4 Standardization

Setting the variance of the state equation, Q, to one gives us mean values for α that lie between -23 and 31. These were normalized such that the expected value for all countries has mean zero and standard deviation one. Each drawn value of $\alpha^{(j)}$ is modified in the following way:

$$BCI^{(j)} = \frac{\alpha_{i,t}^{(j)} - \bar{\alpha}}{\left(\frac{1}{p}\sum_{i=1}^{p}\frac{1}{n}\sum_{t=1}^{n}(\bar{\alpha}_{i,t} - \bar{\alpha})^2\right)^{\frac{1}{2}}}$$

with $\alpha_{i,t}^{(j)}$ the value of α for country i at time t in the j^{th} iteration; $\bar{\alpha}_{i,t}$ is the mean of alpha over all iterations, and $\bar{\alpha}$ is the mean over all iterations, years and countries. The *BCI* index for country i at time t and its variance is then respectively the mean and variance of $\mathrm{BCI}_{i,t}^{(j)}$ with respect to j.

yearly standardization

The normalization used in the WGI on the other hand is such that the mean values for all countries has a *yearly* mean of zero and standard deviation one.

$$BCI^{(j)} = \frac{\alpha_{i,t}^{(j)} - \frac{1}{p}\sum_{i=1}^{p}\bar{\alpha}_{i,t}}{\left(\frac{1}{p}\sum_{i=1}^{p}\left(\bar{\alpha}_{i,t} - \frac{1}{p}\sum_{i=1}^{p}\bar{\alpha}_{i,t}\right)^2\right)^{\frac{1}{2}}}$$

3.A.5 Convergence

The Gibbs sampler algorithm ran 100,000 iterations of which the first 50,000 were discarded as burn-in. To ensure that the model has converged, the draws of the (more than 400) parameters of state and measurement equation were individually examined using simple plotted values, autocorrelation functions and a rolling window CUMSUMs. Figure 3.10 illustrates this for the parameter $Z_(41,1)$. All plots point to a well-behaved, converged distribution, which is what is found for the other parameters as well.

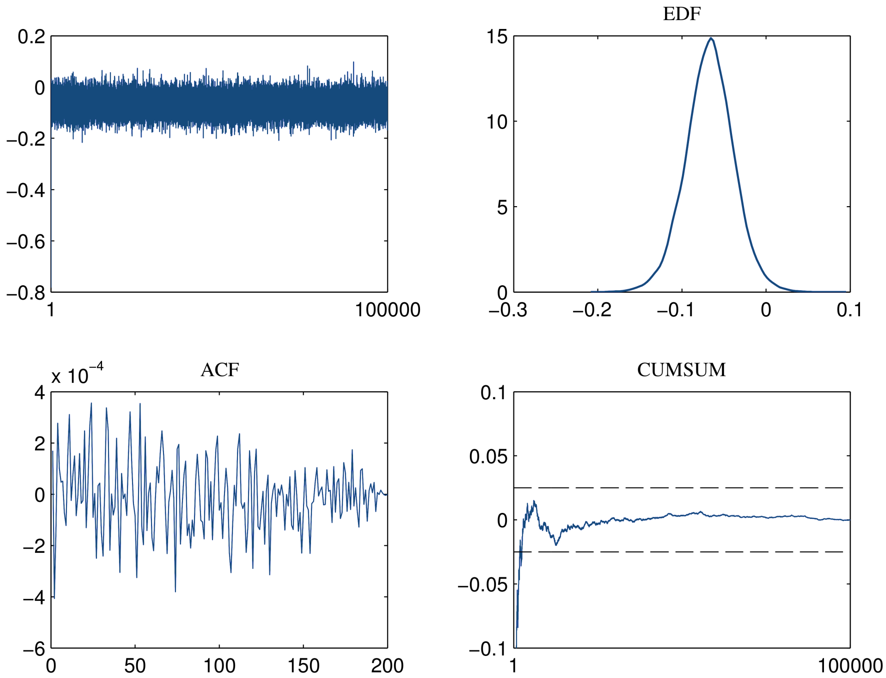


Figure 3.10: Convergence statistics for $Z_{(41,1)}$
Top left: simple plot of all drawn values; top right: the empirical distribution function; bottom left: the autocorrelation function; and bottom right: the rolling window CUMSUM statistic, with 95% significance bounds (window: 1000 draws).

3.A.6 Model selection

Chibbs' method enables the computation of the marginal probability of complex models using any collection of parameter values $\theta^\star$ (Carlin and Louis, 2000):

$$\log \hat{p}(y|M_i) = \log f(y|M_i, \theta^\star) + \log p(\theta^\star|M_i) - \log p(\theta^\star|M_i, y) \tag{3.31}$$

with $\theta^\star = (C^\star, Z^\star, H^\star, T_i^\star)$ for the baseline, cross-correlated and $T_i = \tau$ models, $\theta^\star = (C^\star, Z^\star, H^\star)$ when T is set to zero, and $\theta^\star = (C^\star, Z^\star, D^\star, H^\star, T_i^\star)$ for the model with persistent measurement errors. In our computations, the expected parameter values from the converged models were used as $\theta^\star$.

The last term in equation 3.31 can be split up into a number of conditional probabilities in the following way (suppressing the M_i term to keep notation simple):

$$\begin{aligned}
\log p(C^\star, Z^\star, H^\star|y) &= \log p(C^\star, Z^\star|H^\star, \alpha^\star, y) \\
&\quad + \log p(H^\star|\alpha^\star, y) \\
&\quad + \log p(\alpha^\star|y) \qquad (3.32) \\
\log p(C^\star, Z^\star, H^\star, T_i^\star|y) &= \log p(C^\star, Z^\star|H^\star, T_i^\star, \alpha^\star, y) \\
&\quad + \log p(H^\star|T_i^\star, \alpha^\star, y) \\
&\quad + \log p(T^\star|\alpha^\star, y) \\
&\quad + \log p(\alpha^\star|y) \qquad (3.33) \\
\log p(C^\star, Z^\star, D^\star, H^\star, T_i^\star|y) &= \log p(C^\star, Z^\star|y, D^\star, H^\star, T_i^\star, \alpha^\star, y) \\
&\quad + \log p(D^\star|H^\star, T_i^\star, \alpha^\star, y) \\
&\quad + \log p(H^\star|T_i^\star, \alpha^\star, y) \\
&\quad + \log p(T^\star|\alpha^\star) \\
&\quad + \log p(\alpha^\star|y) \qquad (3.34)
\end{aligned}$$

When conditioning on $\alpha^\star$, the distributions of C, Z, D and H do not depend on T_i, and vice versa, which simplifies the computations significantly. Carlin and Louis (2000, chapter 6.2) explain in detail how to calculate these conditional probabilities, where it should be noted that $p(H^\star|T_i^\star, \alpha^\star, y)$, $p(H^\star|\alpha^\star, y)$ and $p(D^\star|H^\star, T_i^\star, \alpha^\star, y)$ require an additional Gibbs sampler to be run. In addition, $p(\alpha^\star|y)$ is computed using the Kalman smoother. Lastly, the Bayes factor is not defined when using flat priors, which is why the actual (uninformative) priors from 3.A.1 are used, where D has the same prior as T.

3.B Summary of the used corruption indicators

Table 3.9: Summary of the used corruption indicators

Source	Name	Indicator	Obs	Corr. BCI	group
African Development Bank$^{(A)}$	ADB	Transparency, accountability and corruption in public sector	388	0.618	1
Afrobarometer		How many of the following people are involved in corruption:			
	AFR2	... elected leaders	64	-0.4	2
	AFR3	... civil servants	64	-0.533	2
	AFR4	... president	54	-0.591	2
	AFR5	... police	54	-0.72	2
	AFR6	... judges and magistrates	54	-0.701	2
Asian Development Bank$^{(A)}$	ASD	Transparency, accountability and corruption in public sector	192	0.388	3
Business Environment and Enterprise		Frequency of unofficial payments/gifts...			
performance Survey	BPS1	... to deal with occupational health and safety inspections	61	-0.375	4
	BPS2	... to deal with fire and building inspections	61	-0.667	4
	BPS3	... to deal with environmental inspections	61	-0.596	4
	BPS8	... to deal with customs/imports	117	-0.533	4
	BPS12	... to get connected to public services	86	-0.553	4
	BPS13	... to get licenses and permits	86	-0.662	4
	BPS14	... to deal with taxes and tax collection	117	-0.689	4
	BPS15	... when dealing with courts	117	-0.53	4
	BPS19	... to influence the contents of new laws	86	-0.23	4
	BPS30	... obtain government contracts	86	-0.24	4
	BPS10	% of annual sales typically paid in bribes by a firm like yours	117	-0.284	4
	BPS11	How corrupt/honest are court systems?	117	0.182	4
	BPS20	Frequency of informal payments to get things done	117	-0.629	4
	BPS21	% of contract value typically paid as bribe in government contract	117	-0.271	4
	BPS26	Total annual informal payment	31	-0.335	**19**
	BPS27	Firms in my line of business know in advance how much to bribe	86	-0.535	4
Economist Intelligence Unit$^{(A)}$	EIU	Corruption among public officials	1840	0.893	5
Freedom House - Nations in transit	FRH	Control of corruption	382	-0.777	6

Table 3.9: Summary of the used corruption indicators

Source	Name	Indicator	Obs	Corr. BCI	group
Global Corruption Barometer (TI)		Perceptions of corruption for each of the following institutions:			
	GCB1	... political parties	421	-0.466	7
	GCB2	... parliament/ legislature	421	-0.601	7
	GCB10	... police	349	-0.484	7
	GCB11	... business/ private sector	351	-0.695	7
	GCB12	... media	251	-0.821	**28**
	GCB13	... public officials/ civil servants	251	-0.681	**28**
	GCB14	... judiciary	251	-0.59	**28**
	GCB15	... NGOs	250	-0.769	**28**
	GCB16	... religious bodies	131	-0.895	**20**
	GCB3	... military	350	-0.806	7
	GCB4	... education system	421	-0.161	7
	GCB5	... registry and permit	421	0.058	7
	GCB6	... medical services	169	-0.635	**27**
	GCB7	... utilities	420	-0.765	7
	GCB8	... tax revenue	352	-0.223	7
	GCB9	... customs	350	0.184	7
		corruption affects			
	GCB27	... politics	239	-0.455	**26**
	GCB28	... business environment	239	-0.387	**26**
	GCB29	... personal life	239	-0.662	**26**
Global Competitiveness Survey	GCS1	Public trust in politicians	923	0.849	8
	GCS2	Diversion of public funds	923	0.967	8
		How common are bribes in...			
	GCS3	... export-import	1062	0.928	8
	GCS4	... utilities	1062	0.899	8
	GCS5	... tax collection	1062	0.914	8
	GCS6	... public contracts	1062	0.945	8
	GCS7	... judicial decisions	1062	0.936	8
Gallup World Poll	GWP	Is corruption in government widespread?	967	0.634	9
IFAD rural sector performance	IFD	Transparency, accountability and corruption in rural areas	856	0.515	10
Institutional Profiles Database	IPD1	Level of "petty corruption"	257	0.858	11
	IPD2	Level of "large-scale corruption"	257	0.835	11
Latinobarometer	LBO1	The biggest problem in the country is corruption	260	-0.242	12
	LBO3	Evolution of corruption over the past 5 years	68	0.229	**24**
	LBO4	How serious is the corruption problem	51	0.419	**24**
	LBO6	% of corrupt civil servants	54	-0.777	**22**

Table 3.9: Summary of the used corruption indicators

Source	Name	Indicator	Obs	Corr. BCI	group
	LBO8	Probability of bribing a police man	37	-0.7	**25**
	LBO9	Probability of bribing a judge	37	-0.76	**25**
	LBO10	Probability of bribing someone in a ministry	37	-0.677	**21**
Country Policy and Institutional Assessment	PIA	Transparency, accountability and corruption in the public sector	536	0.634	13
Political and Economic Risk Consultancy in Asia$^{(A)}$	PRC	To what extent does corruption exist and hinder business	189	-0.942	14
International Country Risk Guide (PRS)	PRS	Control of corruption	3868	0.766	15
Vanderbilt	VAB1	The most serious problem facing the country is corruption	92	-0.352	16
	VAB4	Level of corruption in political parties	21	0.023	
	VAB5	Level of corruption in public officials	91	0.47	16
World Competitiveness Yearbook	WCY	Bribing and corruption exist in the economy	928	0.956	17
Global insight$^{(A)}$	WMO	Control of corruption	2538	0.903	18

$^{(A)}$Data from WGI: www.govindicators.org

3.C Selection bias in WGI and CPI

Table 3.10: Selection bias regressions: corruption only

					WGI				
	1995	1996	1997	1998	1999	2000	2001	2002	2003
BCI	-	-0.2034*	-	-0.1549*	-	-0.1705*	-	-0.2082*	-0.2148**
		(0.1131)		(0.1226)		(0.1315)		(0.1352)	(0.1304)
constant	-	1.1469***	-	1.3926***	-	1.4631***	-	1.513***	1.5119***
		(0.1104)		(0.1272)		(0.1283)		(0.1363)	(0.1324)
covered		184		194		196		197	197
total		212		212		212		212	212
					CPI				
BCI	0.631***	0.4198***	0.54***	0.2982***	0.216***	0.278***	0.2772***	0.1835**	0.0319
	(0.1109)	(0.0988)	(0.1046)	(0.0909)	(0.0879)	(0.0906)	(0.0919)	(0.0906)	(0.09)
constant	-0.9887***	-0.7047***	-0.771***	-0.2285***	-0.0668	-0.167**	-0.1803**	-0.0445	0.3147***
	(0.1131)	(0.0991)	(0.1059)	(0.0902)	(0.0881)	(0.0879)	(0.0881)	(0.0877)	(0.0877)
covered	42	54	52	87	100	92	91	102	132
total	212	212	212	212	212	212	212	212	212

					WGI				
	2004	2005	2006	2007	2008	2009	2010	2011	2012
BCI	0.0644	0.0813	0.0402	0.0512	0.0376	0.2195	0.1916	0.6012*	-
	(0.1749)	(0.1826)	(0.1697)	(0.1855)	(0.1722)	(0.26)	(0.2507)	(0.513)	
constant	1.8148***	1.8241***	1.8138***	1.8846***	1.8621***	2.1901***	2.1783***	2.8116***	-
	(0.1596)	(0.1687)	(0.167)	(0.1671)	(0.1682)	(0.2311)	(0.2172)	(0.5769)	
covered	204	204	205	206	207	210	210	212	
total	212	212	213	213	214	214	214	214	
					CPI				
BCI	-0.054	-0.2126**	-0.1888**	-0.3307***	-0.3426***	-0.3149***	-0.3582***	-0.3617***	-0.325***
	(0.0913)	(0.096)	(0.0951)	(0.1084)	(0.104)	(0.108)	(0.1056)	(0.1096)	(0.1037)
constant	0.4808***	0.6951***	0.7326***	1.0704***	1.0517***	1.0477***	1.0136***	1.1209***	0.9625***
	(0.0885)	(0.0945)	(0.0965)	(0.1116)	(0.111)	(0.1076)	(0.1083)	(0.1143)	(0.1068)
covered	145	159	163	180	180	180	178	183	176
total	212	212	213	213	214	214	214	214	214

Table 3.11: Selection bias regressions: corruption and GDP

	1995	1996	1997	1998	1999	2000	2001	2002
WGI								
BCI	-	-0.4491**	-	0.0566	-	0.1063	-	-0.0779
		(0.1924)		(0.5201)		(0.3805)		(0.3891)
GDP	-	35168.9907***	-	12360.5732***	-	11719.2989***	-	12357.5924***
		(3241.0577)		(4857.7755)		(2531.0071)		(2858.59)
constant	-	1.086***	-	2.6821***	-	2.1772***	-	2.4057***
		(0.1811)		(0.5188)		(0.3686)		(0.3802)
covered		176		186		186		187
total		187		187		188		188
CPI								
BCI	0.798***	0.2833**	0.5811***	0.1975**	0.1491*	0.2107**	0.2034**	0.0881
	(0.1554)	(0.1434)	(0.1695)	(0.1204)	(0.1132)	(0.1125)	(0.1165)	(0.1122)
GDP	5879.397***	13954.1899***	13718.3696***	11838.4062***	7217.5041***	6618.2486***	7554.3935***	7144.4872***
	(871.9195)	(1714.8077)	(2448.6348)	(2938.9472)	(1490.3057)	(1279.1003)	(1119.5978)	(1853.3881)
constant	-1.7037***	-1.6435***	-1.8437***	-0.7188***	-0.3494***	-0.4824***	-0.5628***	-0.3832***
	(0.1922)	(0.1727)	(0.2318)	(0.1529)	(0.121)	(0.1212)	(0.1179)	(0.1381)
covered	42	54	52	87	100	92	91	102
total	187	187	187	187	187	188	188	188

	2003	2004	2005	2006	2007	2008	2009	2010
WGI								
BCI	-0.0113	-0.0207	-0.0636	-0.0469	-0.1001	-	-	-
	(0.4713)	(0.5155)	(0.5163)	(0.5025)	(0.4424)			
GDP	4754.0892***	9692.2166***	19955.0621***	9557.5311***	1483.3758*	-	-	-
	(1469.0676)	(1461.2844)	(2001.7873)	(2579.0869)	(1087.9154)			
constant	2.6186***	2.5595***	2.4082***	2.5945***	2.7742***	-	-	-
	(0.4183)	(0.4436)	(0.5336)	(0.4572)	(0.5152)			
covered	187	187	187	188	188			
total	188	188	188	189	189			
CPI								
BCI	-0.1095	-0.2001*	-0.4538***	-0.3438**	-0.6221***	-0.687***	-0.505***	-0.547***
	(0.1422)	(0.1443)	(0.1501)	(0.1505)	(0.1928)	(0.2046)	-0.1961	(0.1634)
GDP	30739.2816***	25916.142***	24618.4656***	14618.7394***	9643.676***	9340.673***	7462.4153***	5650.3246***
	(4650.8969)	(2968.2047)	(2423.7507)	(3191.1543)	(1540.5052)	(1451.8276)	-1875.9244	(1400.4437)
constant	-0.4088***	-0.0933	0.1684	0.4737***	1.285***	1.3319***	1.328***	1.2093***
	(0.1502)	(0.1365)	(0.1388)	(0.148)	(0.1868)	(0.2013)	-0.1929	(0.1759)
covered	130	143	156	162	179	179	179	179
total	188	188	188	189	189	189	189	189

3.D Changes in the block-diagonal BCI

Table 3.12: Changes in the level of the block-diagonal BCI between 2000 to 2010

	Deteriorated (BCI decreased)		Improved (BCI increased)	
1%	Greece^{WGI}		Georgia^{WGI}	Qatar^{WGI}
	Italy^{WGI}		Macedonia^{WGI}	Rwanda^{WGI}
5%	Fiji	Senegal	Algeria	Paraguay
	Great Britain^{WGI}	South Africa	Bangladesh	Saudi Arabia
	Kuwait	Turkmenistan	Indonesia	Serbia^{WGI}
	Madagascar		Japan	Slovenia
	Malaysia		Lesotho	Turkey
	Moldova		Liberia	
10%	Brazil	USA	Congo, Dem. Rep.	Latvia
	Burkina Faso		Cuba	Palestine
	Hungary		Djibouti	Serbia and
	Iran		Ghana	Montenegro
	Papua New Guinea		Hong Kong	UAE^{WGI}
	Philippines		Iraq	Uruguay
	Trinidad and Tobago		Jordan	Zambia
not	Eritrea^{WGI}			
	Venezuela^{WGI}			

List of the countries whose level of corruption changed significantly between 2000 and 2010 according to the cross-correlated BCI. ^{WGI} indicates whether the change was detected using WGI.

3.E Variance decomposition

Table 3.13: Variance decomposition and goodness of fit of the measurement equation

	$R^2_{overall}$	$R^2_{between}$	R^2_{within}		$R^2_{overall}$	$R^2_{between}$	R^2_{within}
GCS2	0.924	0.945	0.538	GCB6	0.385	0.378	0.095
WCY	0.916	0.96	0.165	ADB	0.383	0.418	0.012
PRC	0.875	0.932	0.083	GCB2	0.351	0.243	0.084
GCS6	0.873	0.929	0.561	GCB14	0.344	0.348	0.001
GCS7	0.869	0.902	0.492	BPS3	0.329	0.475	0.397
GCS3	0.843	0.912	0.499	AFR3	0.308	0.42	0.003
GCS5	0.827	0.858	0.5	AFR4	0.306	0.372	0.186
WMO	0.821	0.876	0.056	BPS27	0.285	0.659	0.341
EIU	0.802	0.854	0.031	BPS12	0.282	0.389	0.29
GCS4	0.8	0.834	0.413	BPS8	0.282	0.513	0.032
GCB16	0.791	0.782	0.085	BPS15	0.274	0.446	0.023
IPD1	0.742	0.769	0.003	IFD	0.272	0.343	0.002
GCS1	0.71	0.733	0.199	GCB10	0.226	0.203	0.039
FRH	0.702	0.758	0.072	VAB5	0.215	0.43	0.007
IPD2	0.695	0.758	0.002	GCB1	0.21	0.11	0.058
GCB12	0.668	0.672	0.013	GCB27	0.201	0.156	0.023
GCB3	0.633	0.565	0.068	LBO4	0.188	0.314	0.003
LBO6	0.624	0.725	0	AFR2	0.17	0.243	0.017
GCB15	0.58	0.608	0	GCB28	0.144	0.153	0.007
GCB7	0.579	0.419	0.033	ASD	0.124	0.207	0.034
LBO9	0.56	0.658	0.001	BPS1	0.121	0.215	0.333
PRS	0.545	0.758	0.105	VAB1	0.115	0.266	0.004
AFR5	0.514	0.538	0.005	BPS10	0.086	0.365	0.083
AFR6	0.495	0.531	0.065	BPS26	0.082	0.082	–
BPS14	0.488	0.527	0.06	LBO1	0.079	0.364	0.037
GCB11	0.475	0.479	0.013	BPS21	0.072	0.248	0.026
GCB13	0.462	0.47	0	BPS11	0.065	0.176	0.086
LBO8	0.457	0.538	0.011	LBO3	0.062	0.149	0.001
BPS2	0.443	0.556	0.46	GCB8	0.046	0.013	0.014
LBO10	0.431	0.517	0.017	GCB9	0.04	0.087	0.068
GCB29	0.428	0.414	0.01	BPS30	0.037	0.149	0.039
BPS20	0.415	0.648	0.01	BPS19	0.034	0.077	0.092
BPS13	0.401	0.596	0.212	GCB4	0.024	0	0.031
PIA	0.396	0.427	0.017	GCB5	0.004	0.009	0.001
GWP	0.39	0.412	0.129	VAB4	0.002	0.002	–

List of the R^2 of the measurement equation for each indicator of corruption, in descending order of their goodness of fit. The last two columns decompose the overall R^2 into its between and within components (Verbeek, 2010).

4 | Measuring Actual Economic Integration - An outline of a Bayesian state-space approach[1]

Abstract

In spite of the spectacular growth of integration agreements, there does not exist a standard, systematic index that measures regional integration. The most likely reason for this dearth is that it is hard to capture this complex and multidimensional process in a single indicator. Even the most basic definition of regional integration encompasses many different aspects, increasing the difficulty of finding appropriate data exponentially. This chapter proposes using the state-space model to construct an index of the level of integration. The versatility of this approach allows us to capture the multidimensional nature of integration, even when using data of varying quality and availability. We build a bilateral index of the level of Actual Economic Integration, capturing trade in goods and services, migration and financial flows. Gravity model estimations reveal that while the EU and Nafta were closed between countries that were already highly integrated, these agreements succeeded in further raising the level of economic integration both in the short and long term.

Keywords: Actual economic integration; State-space model; Gravity equation;

[1]This chapter is the result of joint work together with Prof. Dr. Glenn Rayp.

Free trade agreements.

JEL: F15; F55; C43.

4.1 Introduction

Despite widespread academic and policy interest and in contrast with other aspects of institutional and international economics (e.g. governance or globalization) there does not exist a standard, systematic index of regional integration. Such a measure could be used to determine the trends in the world economy more precisely (e.g. the link between globalization and regionalization), to monitor integration policy initiatives more accurately and to assess the effectiveness of current or past policy initiatives (e.g. indicating good practices). Nevertheless, in their review, De Lombaerde et al. (2008, p.2) note that *'only a few attempts have been undertaken to design composite indices of regional integration and no proposal has been systematically and continuously used as a policy tool.'*

The most plausible explanations for this lack are data availability and methodological issues. Regional integration is a complex and multidimensional process and therefore difficult to capture in a single or even a few indicators. Typically, a larger set of indicators is used, usually of very different quality in which scoring by the analyst is not uncommon.

Studies of the efficacy of regional integration agreements such as the EU or Nafta have for the most part focused on the effect on bilateral trade flows. However, regional integration agreements typically also include provisions dealing with trade in services, migration and financial flows. Moreover, the determinants of these bilateral flows are analyzed using the same estimation framework: the gravity model. This means that an index of the overall level of integration would allow us to combine these analyses into a single study that matches the scope of the agreements.

Constructing an indicator that integrates the information of all the available data, immediately brings up the question of how to summarize (i.e. aggregate) the individual indicators and which weighting scheme to use. For example, Feng and

Genna (2003) follow Hufbauer and Scott (1994) in their construction of Integration Achievement Scores by using the simple arithmetic average of the categories that measure distinct components of (institutional) regional integration. The index of institutional regional integration in Dorrucci et al. (2004) is also computed as an unweighted average of assigned achievement scores in each of the Balassa stages in regional integration. This index is subsequently linked to a set of indicators of actual economic integration in order to study causal effects. In UNECA (2001, 2002, 2004) the composite index is constructed as a weighted mean: first at the country level taking expert opinions as the basis of the weighting scheme; Second at the regional level, using country GDPs as weights. Dennis and Yusof (2003) take as composite integration indicator the simple arithmetic average of a small subset of their key indicators. Finally, the UN-ESCWA (2006) report uses a principal component analysis to compute the level of actual integration of Arab countries.

In this contribution, we propose a new approach to constructing a regional integration indicator, that is a bayesian state-space approach, which can circumvent the data quality and availability issues allowing a systematic and continuous use.

De Lombaerde et al. (2011) formulate a three-step method in constructing a composite index. The first step concerns the principles on which the individual indicators of the index should be based: relevance, accuracy and credibility, data availability, timeliness and comparability. Often, these principles are (partially) neglected out of necessity: the lack of indicators that take the multidimensionality of regional integration into account compels to use incomplete or inaccurate data. Of course, this is common to whichever method is used to construct an aggregate indicator. However, in contrast to other methods that have been used, the state-space approach can take the uncertainty of the data into account as well as correct for missing values in a statistically transparent way.

The second step of De Lombaerde et al. (2011) refers to the classification of the variables according to particular aspects of regional integration, e.g. the distinction between indicators of the actual integration process and the institutional characteristics. The state-space approach allows for such a functional distinction between the indicators and can deal with this in two ways: either as separate composite indices, which can be further used for analytical purposes, or as components of a more general index, in which case their respective weights are informative about the impact on the integration process just like their correlation gives an indication of their complementarity.

The third and final step of De Lombaerde et al. (2011) consists of the construction of the composite regional integration index, in particular the issues of the determination of the weighting scheme for the indicators (e.g. statistical or not) and the method of aggregation (e.g. arithmetic mean or more involved). There, the bayesian state-space approach offers the advantage of making fewer assumptions in determining the indicators' weights and of being more transparent in the aggregation.

In the next section, we describe the principles of the bayesian state-space methodology. We keep this description brief and refer the interested reader to more formal and thorough treatments of the subject. To show the potential of the approach, the third section discusses the construction of an indicator of actual economic integration at the bilateral country level for the member countries of the OECD. For these countries a sufficiently large amount of data is available that due account can be taken of the multidimensionality of regional integration. Based on this first application, we consider in the fifth section the extensions that the method allows. Conclusions are drawn in the last section.

4.2 Methodology

This section only aims to give a limited overview of the state-space methodology. For more information on state-space models and how to estimate them, see Kim and Nelson (1999) or Durbin and Koopman (2012). More detailed information on this particular model can also be found in Standaert (2014) where it is used to combine indicators of corruption.

4.2.1 The state-space model

The main idea in the state-space model is to estimate the unknown overall level of regional integration *RI* (the *state* variable), using the information in the different indicators of regional integration, y.[2] In order to understand how this happens, it is necessary to go back to the two equations that define its workings: the measurement (4.1) and state equation (4.2).

$$
\begin{aligned}
y_{i,t} &= C + Z * RI_{i,t} + \varepsilon_{i,t} && (4.1)\\
RI_{i,t} &= T_i * RI_{i,t-1} + \nu_{i,t} && (4.2)\\
\varepsilon_{i,t} &\sim N(0,H) && (4.3)\\
\nu_{i,t} &\sim N(0,1) && (4.4)
\end{aligned}
$$

for all country-couples $i \in [1,n]$ and years $t \in [1,N]$.

The measurement equation states that the k indicators of regional integration $y_{i,t}$ (e.g. the level of bilateral trade) depend on the overall level of integration $RI_{i,t}$. The error term $\varepsilon_{i,t}$ captures differences in the quality of the indicators whether due to measurement errors or because of the influence of factors other than the level of integration. The better an indicator $y^{(j)}$ measures the level of integration, the smaller the variance of the corresponding error term $H_{(j,j)}$.

The $(1 \times k)$ vectors C and Z rescale the indicator variables to put them on equal footing. The intercept also controls for the possibility that an indicator covers only the relatively higher or lower integrated, or that it consistently over- or underestimates the level of integration. For example, a standardized index that covers relatively more integrated countries would receive a negative intercept. The exact rescaling parameters are indicator-specific, but are kept constant over time, t, and country-couples, i. Similarly, all indicators can differ in terms of their reliability (H), but

[2] For the sake of readability, the notation is sometimes simplified. $y^{(j)}$ is a single indicator of integration, j, for all country-couples and all years. $y_{i,t}$ is the vector of all indicators in a given year, t, and for a given country-couple, i, while this vector for all years and all country-couples is abbreviated to y.

the reliability of an indicator does not change over time or country-couples.

The state equation (equation 4.2) allows the current level of integration to depend on its past values. The level of dependence, T_i, can vary for each country-couple, i, but is restricted to lie within the [-1,1] interval. This rules out ever-increasing values for the RI index and ensures that the model converges to a steady solution. It does not preclude non-stationarity in the level of regional integration.

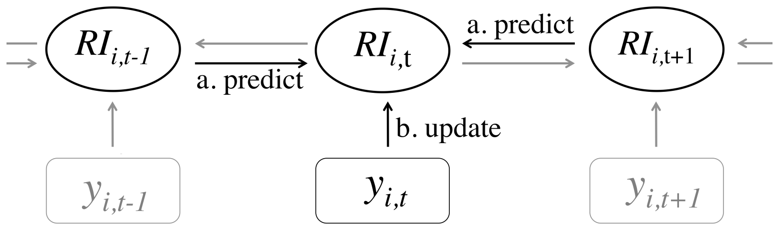


Figure 4.1: Estimation using time dependency

Figure 4.1 illustrates the advantage of adding the time-dependency in the state equation. To the extent that the level of integration depends on its previous values, both past and future information is used to predict what the level of integration is at time t (step a). This prediction is governed by the state equation (equation 4.2): the weaker the link between AEI and its preceding values, the more uncertain the prediction becomes. This forecasted value is subsequently compared to the indicators of integration today, y_t, using the parameter values in the measurement equation (equation 4.1). If the y_t contains new information, the estimated level of integration is adjusted (step b). The more reliable an indicator is, the bigger the influence of step b is.

Because the $RI_{i,t-1}$ and $RI_{i,t+1}$ in turn depend on their past values and future values, the entire time-series is used when estimating the current level of integration. The advantages are manifold. First of all, it significantly increases the number of years for which the indicator can be reliably computed. Moreover, the increase in information helps the algorithm to better distinguish between random measurement errors and the actual changes in the level of integration. This results in smoother estimates made with smaller confidence bounds.

The main strength of the state-space model is the ease with which it handles missing observations. Simply put, missing observations are replaced by information which has absolutely no value: $y_{i,t}^{(k)} = 0$ and $\text{var}(\varepsilon_{i,t}^{(k)}) = \infty$. This allows the model to run uninterruptedly without fundamentally changing the value of missing data. Because the entire time-series is used when estimating the value of *RI*, this negates the need to impute or otherwise manipulate missing data ex ante (Kim and Nelson, 1999; Durbin and Koopman, 2012).

An additional advantage of this model is that it encapsulates a number of other techniques. For example, if we assume that *RI* does not depend on its previous values ($T = 0$) and all indicators have the same reliability ($H_{(j,j)} = c_H$), it can be shown that this model will return a principal component analysis. If in addition it is assumed that all indicators are scaled the same way ($Z_{(j,1)} = c_Z$ and $C_{(j,1)} = c_C$), then it returns a simple average.

In other words, the usefulness of the state-space approach follows directly from the validity of the assumptions on the parameter values. If the level of integration is not expected to depend on its previous values ($T = 0$) a principal component analysis will suffice. However, if these assumptions are incorrect, using more simple techniques discards information and could lead to incorrect conclusions. It also means that it is possible to test the validity of the state-space model using its parameter values.

4.2.2 Bayesian estimation

In order to estimate the state-space model, it is necessary to solve for $RI_{i,t}$ as well as the parameters of the state and measurement equation: H, Z, C and T. As the number of countries and years increases, this estimation becomes increasingly cumbersome. However, using a bayesian Gibbs sampler, it can be split up into different sections that can be dealt with one at a time.

If the values for $RI_{i,t}$ were known, the state and measurement equations would be simple linear regressions and we could easily compute and draw from their distributions. Similarly, if the parameters were known, we could draw from the distribution of *RI* using a simulation smoother (Durbin and Koopman, 2012). It can be shown that by iteratively drawing from both distributions while conditioning on the last drawn value, these draws will converge to the unconditional distribution. After discarding the first non-converged values (the burn-in), the remaining drawn values can be used to reconstitute the original unconditional distribution of *RI* as well as those of the parameters. For more information on Bayesian econometrics and Gibbs sampling see Lancaster (2004) and Koop et al. (2007).

Because this model is estimated in a Bayesian framework, it is necessary to be explicit about the prior distribution of the parameters. However, seeing that there is no ex-ante information on them we use flat priors. This means that these parameters are not restricted in any way. The only variables that are limited are T_i, whose values have to lie inside the [-1,1] interval, and the diagonal elements of the variance H, whose values have to be strictly positive.

$$p(C) \sim \mathbf{1_{(1\times k)}} \tag{4.5}$$

$$p(Z) \sim \mathbf{1_{(1\times k)}} \tag{4.6}$$

$$p(log(H_{(j,j)})) \sim 1 \quad \forall j \in [1,k] \tag{4.7}$$

$$p(T_i) = 0.5 * \mathbb{1}_{[-1,1]}, \quad \forall i \in [1,n] \tag{4.8}$$

4.2.3 Rescaling to a ratio variable

As it is specified in equations 4.1–4.4, the state-space model returns an interval variable. Lacking a fixed zero point, the index can still be compared over time and countries, but ratios of the index would be meaningless. For example, we can say that a country has become more/less integrated, but not that that its level of integration has doubled/halved (cf. temperature in degrees Celsius vs. Fahrenheit).

The reason why the underlying indicators, even when they are ratio variables, are translated into an index that is an interval variable is the intercept in the measurement equations. Without it, the index would be zero if all the underlying variables are zero. However, simply setting $C = 0$ might cause other problems in the state-space model as this parameter also controls for persistent over- and underestimations and selection bias problems.

The solution is to leave the intercepts in the model, but rescale the index afterwards to ensure that the expected value of $RI_{i,t}$ is zero when all underlying indicators are zero. In each iteration of the Gibbs sampler, the expected value of the index is computed when $y = 0$ (conditional on the parameters of the state-space model C, Z, H and T) and this is subtracted from the index. This rescaled index is a ratio variable and will also double when all underlying indicators double making it much easier to interpret.

4.3 An application to the OECD

4.3.1 Defining integration

This section illustrates the state-space approach by measuring the level of bilateral integration between the members of the OECD. Specifically, it examines the level of Actual Economic Integration (AEI) defined by Mongelli et al. (2005, p.6) as *'the degree of interpenetration of economic activity among two or more countries belonging to the same geographic area as measured at a given point in time.'*

This definition is relatively narrow and puts strict limits on the variables to be included. It excludes institutional or cultural integration and even within the concept of economic integration it focuses on actual *interpenetration of activities*. This implies that strictly speaking the co-movement of prices and GDP and other factors from the optimal-currency-area theory should not be included. In addition, it focuses on actual integration as opposed to measuring the potential benefits of integration (e.g. differences in factor endowments).

As a result, the RI indicator computed here is relatively neutral. It does not rely on any specific (economic) theory on integration, nor does it treat integration as necessarily good or bad. It simply measures the extent to which the economic flows of two countries are intertwined. Needless to say, this definition is but one of the many possible choices and the state-space methodology can be easily expanded to include different definitions of regional integration.

The unit of analysis in this study is country couples, and their integration is measured in a directional sense. In other words, the values of the index $AEI_{A,B}$ express to what extent the bilateral economic flows between countries A and B are important for country A. Allowing the values of $AEI_{B,A}$ to differ from $AEI_{A,B}$ makes sense in a network where country size varies significantly. For example, that the German-Estonian trade is important for Estonia does not necessarily imply that the same holds for Germany.

It is important to note that even using Mongelli's definition of integration, many other units of analysis are possible. For instance it would also be possible to study the level of integration of a country within a region. The choice of unit should be primarily driven by the intended use of the indicator. The index of bilateral integration defined here enables us to build a directed and weighted network of integrated countries or use the indicator in a gravity model, but other uses require different definitions and measurement units.

4.3.2 Data

Actual economic integration is measured on four levels, organized according to the 'four freedoms' of the European Single Market: flows of goods, flows of services, FDI and other financial flows and migration. For each a distinction is made between incoming and outgoing flows. Table 4.1 lists the different categories of economic flows in more detail.

Table 4.1: Categories of integration variables

Goods	**Services**	**Migration**	**Financial flows**
- Primary goods	- Total flows	- Foreign population	- Foreign Direct Investment
- Manufactured goods		- Seasonal workers	- Equity
		- Foreign workers	- Debt securities

In order to compare the importance of the flows over countries, these flows are corrected for the size of the sender countries. Specifically, they are normalized both using GDP (population in the case of migration) and as a percentage of total flows. This means that for each category four different variables are used: incoming flows to GDP, incoming flows to total flows, outgoing flows to GDP and outgoing flows to total outgoing flows. The idea is that there are two dimensions in which the bilateral flows can matter for a country: either it covers a significant fraction of total flows and/or it represents a large proportion of GDP. By rescaling the indicators in this way, the size of the country is abstracted from the index: only the relative size of the flows matters.

The bilateral trade data comes from UN Comtrade database. Trade data was collected on the 1 digit level of SITC (v3) product categories. These were subsequently combined into primary goods (categories 1 + 2 + 4 + 68) and manufactured goods (categories 5 + 6 - 68 + 7 + 8) (WTO, 2011, p. 206). The financial flow data on equity and debt securities comes from the IMF's Coordinated Portfolio Investment Survey (CPIS). All other bilateral flow data was gathered form the OECD statistical compendium and the OECD iLibrary. Finally, the data on GDP and population is taken from the Penn World Tables (Heston et al., 2012).

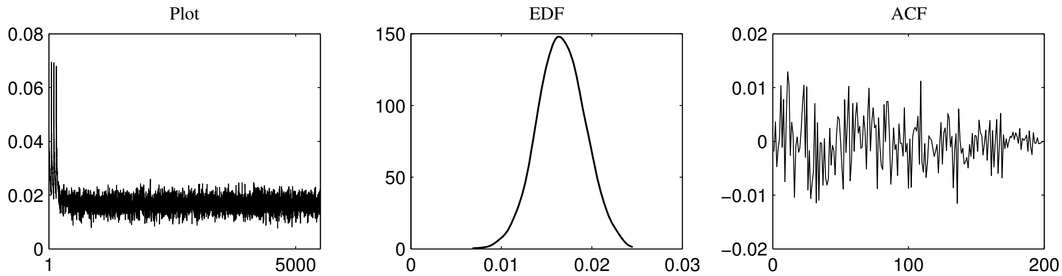
Because of the present lack of detailed data on the flows between non-OECD countries, the dataset was constructed using the 34 current members of the OECD as sender countries. This means that it covers all flows originating from OECD countries, including those directed to non-OECD countries. However indicators of the level of integration from the perspective of non-OECD countries are left out. The target countries were not limited to states, but also include territories (e.g. the French Southern and Antarctic Territories) forming a total of 253 distinct target countries. Dropping those on which no information was available, the dataset contains 8,526 country couples from 1985 to 2012 for a total of 191,493 observations.[3]

The Gibbs sampler ran for 6,000 iterations, of which the first 4,000 were discarded

[3] In order to run the computations on a dataset of this size, we used the resources of the Flemish Supercomputer Center, which was kindly provided by Ghent University, the Flemish Supercomputer Center (VSC), the Hercules Foundation and the Flemish Government - department EWI.

as burn-in. By way of example, figure 4.2 shows the convergence statistics for the scale parameter Z of the outgoing flows of debt securities to GDP. They point to a well-converged posterior distribution, which is what is found for the other parameters in the model as well (available upon request).

Figure 4.2: Convergence statistics of the outgoing flows of debt securities to GDP



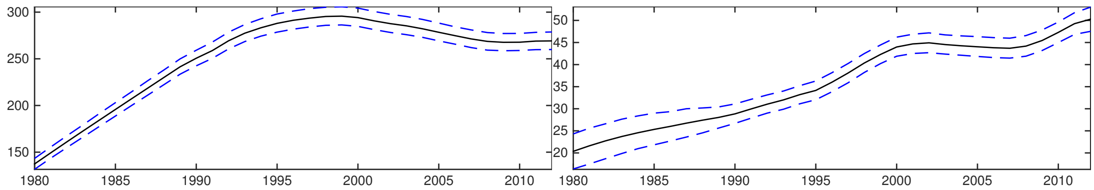
From left to right a simple plot of all the iterated values (including the burn-in) and the empirical distribution function and autocorrelation function of the converged parameter.

4.3.3 An index of Actual Economic Integration

The model returns values for AEI which lie between 0.19 and 296. However, the exact scaling of the AEI index is arbitrary and can be adjusted as long as the relative differences and the zero-point remain unchanged.

By way of illustration, figure 4.3 plots the level of integration and the 90% confidence bounds for two country-couples: USA-Mexico and Mexico-USA. The bilateral flows between Mexico and the United States of America are crucial for Mexico as they lie entirely within the top 99^{th} percentile of all AEI values. The level of integration of the USA in Mexico on the other hand is much lower. Nevertheless, it is both significant and within the top 95^{th} percentile of all values of the index. Mexico's integration into the USA doubles between 1980 and 1995, after which it levels off. In contrast, the USA increases its integration in Mexico throughout the entire time period. The difference in their patterns is probably caused by the increased integration of Mexico in the Chinese economy which rises from less than 5 before 2001 to 24 in 2012.

Secondly, we see that as the number of missing observations decreases, the uncer-



(a) Mexico-USA **(b)** USA-Mexico

Figure 4.3: Plot of normalized AEI indicator with 90% confidence interval (dotted lines)

tainty bounds grow tighter. This is especially clear in the second panel, where the average values are smaller. The number of available indicators is less than 10 for the first 4 years, after which it jumps to 16 in 1989 and keeps rising steadily to 32 in 2008. However, because the entire time series is used in the estimation of each data point, the change in the confidence intervals is much less abrupt than the changes in availability.

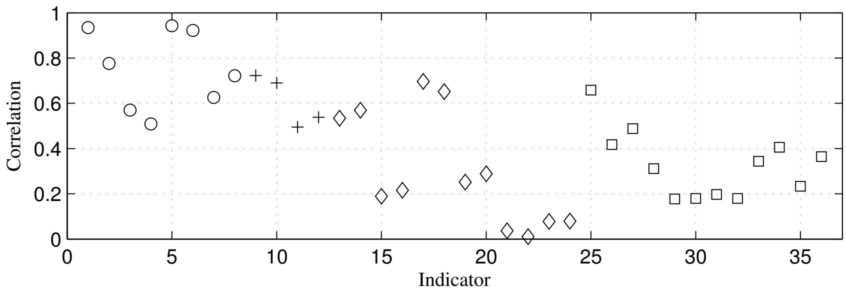


Figure 4.4: Correlation of AEI with individual indicators
Plot of the correlation of the AEI index with the individual indicators from which it was computed.
○ denote trade in goods; + trade in services; ⋄ financial flows; and □ migration.

Returning to the individual indicators of integration, there are only 50 observations where all indicators are available, making it impossible to compare the contribution of all indicators simultaneously. Instead, figure 4.4 graphs their pairwise correlation with the AEI index. It shows that the indicators from all four types of flows are highly correlated with the index, but that migration and the financial flows (represented by diamonds and squares) score relatively worse. A possible explanation is a non-monotonic interaction between the indicators of integration. For example, economic migration is expected to subside when high trade between countries creates opportunities in the home country. On the other hand, the existence of many migrants from a particular country increases the information on potential beneficial trade, leading to a positive correlation. Similarly, having good trading relations increase information which can lead to investment opportunities, but FDI can also be used to circumvent barriers to trade. When both effects are in play simultaneously, the total correlation between the different flows can get muddled. Appendix 4.A lists the R^2 of the measurement equation for each of the indicators of integration, in decreasing order of their goodness of fit.[4] The imports and exports (manufactured and primary) goods and services are found to be the highest correlated with the overall level of integration. The last two columns further decompose the R^2 into its between and within components (Verbeek, 2010), revealing that the between R^2 is markedly higher than the within R^2 for all indicators of integration.

Weighted directed network

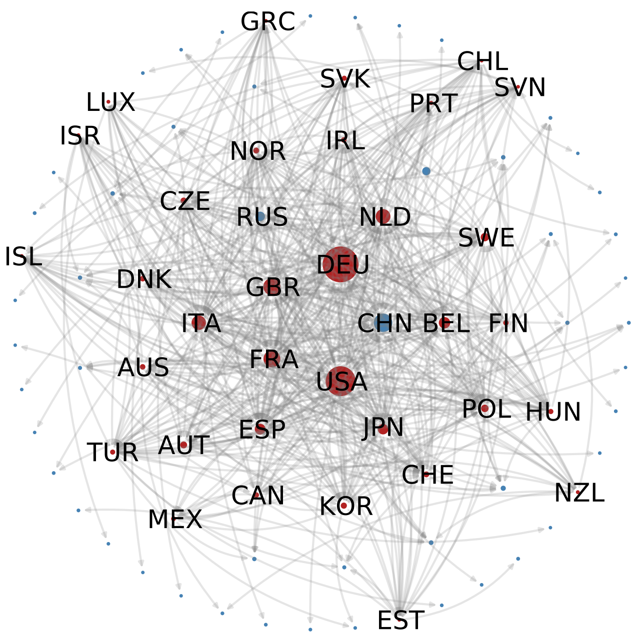


Figure 4.5: Position of the OECD countries in the AEI network in 2010
The nodes represent countries. The higher the weighted indegree, the bigger the node and the closer toward the middle it lies. Nodes are colored red when the country is a member of the OECD.

As mentioned, we can use the values for AEI to construct a network of integration

[4] As these are univariate regressions, the R^2 of this regression is equal to the square of the correlation between AEI and the indicators.

worldwide. Countries are represented by circles (nodes). When one country is highly integrated into another, an arrow (or edge) is drawn between them. The AEI index values represent a natural way of assigning weights to these edges, after resizing them to be strictly positive.

In order to make the network more informative, edges between countries that are not or barely integrated are removed from the network. Instead of choosing an arbitrary cut-off value for the AEI index, the uncertainty of the index values is used to determine the edges. To start, we computed the distribution of the index when all underlying indicators are zero for the entire period and label these values $AEI_{0,t}$. An edge is then said to exist between two countries, if their level of integration is significantly different from $AEI_{0,t}$ at the 1% significance level: $e_{i,t} = 1 \iff AEI_{0,t} < AEI_{i,t}$ in at least 99% of all iterations of the converged Gibbs sampler. This definition identifies 17,102 edges or 9% of all observations.

Figure 4.9 shows the shape of this network in 2010.[5] The values of the index are reflected in the darkness of the arrow between the countries. The position and color of the countries are determined by its weighted indegree: the more important a country is for its partner countries, the darker its color and the more central its position.[6] Countries that are a member of the OECD are indicated in red. Unsurprisingly, it shows that the most central players in the OECD are the USA and Germany, followed by Great Britain, the Netherlands and France. A few non-OECD countries also end up with a high indegree. The importance of China and Russia is driven by intensive trade flows from the OECD countries. Figure 4.6 shows that the distribution of China's (panel a) and Russia's incoming edges (panel b) is similar to that of OECD countries like for example Belgium (panel c).

Density and transitivity

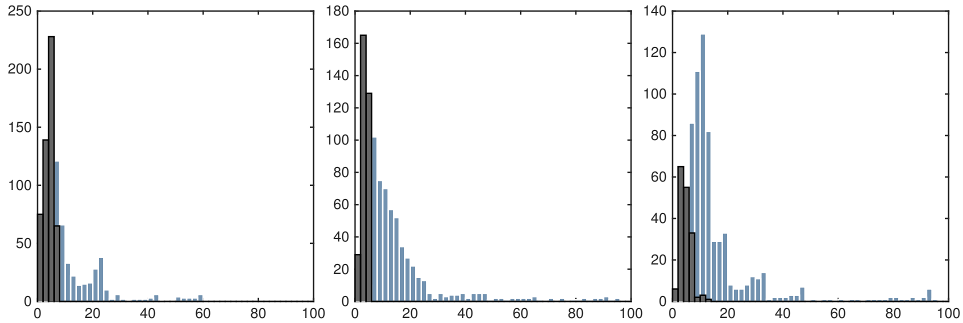
Figure 4.9 gives an overview of the evolution of the network over time. As before, the position of a country on the concentric circles is determined by its indegree. It

[5]The network graphs were generated using the Python package NetworkX

[6]The weighted indegree of a country is the sum of the AEI index of all incoming edges.

Figure 4.6: Incoming-edge distribution of central countries

(a) Russia **(b)** China **(c)** Belgium



Histogram of the incoming edge weights of Russia, China and Belgium. Significant bilateral links (edge = 1) are colored blue, while insignificant links (edge = 0) are colored black

shows that the core of the network stays relatively constant over time. Germany and the USA are by far the most central players in the network followed by France, Great Britain and Italy. Japan quickly rises to a central position in the late 1980s, but then enters a slow decline while still staying in the top ten. Russia and China on the other hand start in the periphery of the network and steadily gain importance over time.

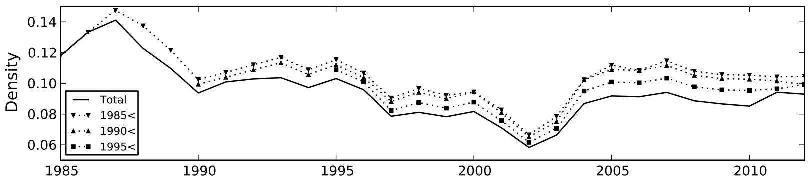
For the most part, the countries in the outermost circle are not members of the OECD, which explains in part their low indegree. There are a few OECD countries that are on the fringes of the network. Ireland for example starts in the periphery and rapidly gains importance in the last decade. However, most peripheral countries do not change position over time (for example Greece, Chile, Slovenia and New Zealand).

While the number of edges (arrows) increases strongly in the first ten years, the plot of the network density[7] (figure 4.7, panel a) reveals the network is becoming less dense over time until 2003, after which this trend is reversed. As the dotted lines show, this pattern is not caused by changes in the number of countries covered

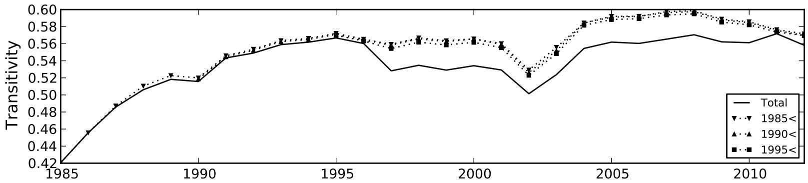
[7] The network density is the fraction of the number of edges between countries divided by the total possible number, measuring the overall connectedness of the network. The network transitivity rates the degree of clustering in the network by looking for triangles: the extent to which a link between countries A-B and A-C also implies that countries B-C are connected. Because these network characteristics are sensitive to changes in data availability, each statistic is recomputed using only the countries that are continuously in the dataset since 1985, 1990 and 1995, respectively.

in the dataset. The network transitivity on the other hand increases strongly over this period. While it is not caused by an increase in the number of countries in the dataset, this initial increase could be driven by the stark increase in the data availability (panel c).

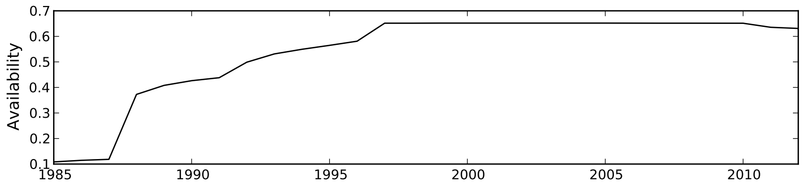
Overall the integration network is characterized by a high transitivity, especially given the density of the network. This pattern corresponds to the strong core-periphery structure visible in figure 4.5 in which there are a few strongly connected countries (the core) with a large periphery feeding into it.[8] Nevertheless, part of the reason why this structure is so prevalent is the fact that links between non-OECD countries are not covered. Except for a dip in 2003, the network density remains relatively constant over time, while the transitivity keeps increasing steadily.



(a) Network density



(b) Network transitivity



(c) Data availability

Figure 4.7: AEI network characteristics over time

[8]Most links are formed in the strongly connected core, resulting in a lot of triangles. Links in the periphery, where such triangles are less likely to form, are more sparse.

4.3.4 Comparison with other techniques

In this section, the AEI index is compared to the two other techniques used in the literature: the simple average and a principal component analysis (pca). Key issue when making this comparison is how to deal with missing observations, since there are only 50 observations where all data is available. For this reason, the mean is calculated over all available values. The weights of the pca are computed using the pairwise correlation matrix, and the index is composed when at least one observation is available.

As was already mentioned, both techniques can be seen as simplified versions of the AEI index where the values of the parameters are restricted in some way. This means that we can test the statistical validity of the state-space approach using the parameter values we find. In this instance, those parameter values confirm the validity of the state-space approach. For example, more than 96% of the time-dependency variables, T_i, are significantly greater than zero at the 1% level. Additionally, the scale parameters are also significantly different over the different indicators, as figure 4.8 shows for the slope parameter Z.

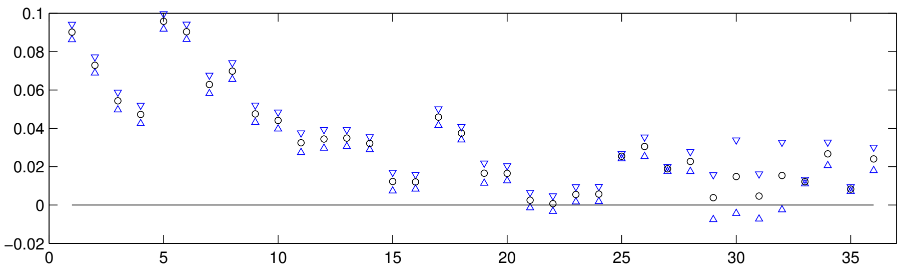


Figure 4.8: Plot of the mean (○) and 95% confidence interval (△) of the scaling parameter Z

The correlation (table 4.2) shows that the result can significantly differ depending on which technique is used: the overall correlation with the adjusted pca is only 0.78 and the adjusted mean scores even worse with 0.53. The high correlation between the three techniques when there are no missing observations suggests that the different solutions to the problem of missing variables is what drives these differences.

Table 4.2: Correlation with mean and principal component analysis

	Obs.	Overall	Between[(a)]	Within[(b)]
mean	50	0.972	0.966	0.040
pca	50	0.961	0.970	−0.021
mean (adj.)	179242	0.534	0.754	0.361
pca (adj.)	179242	0.779	0.927	0.293

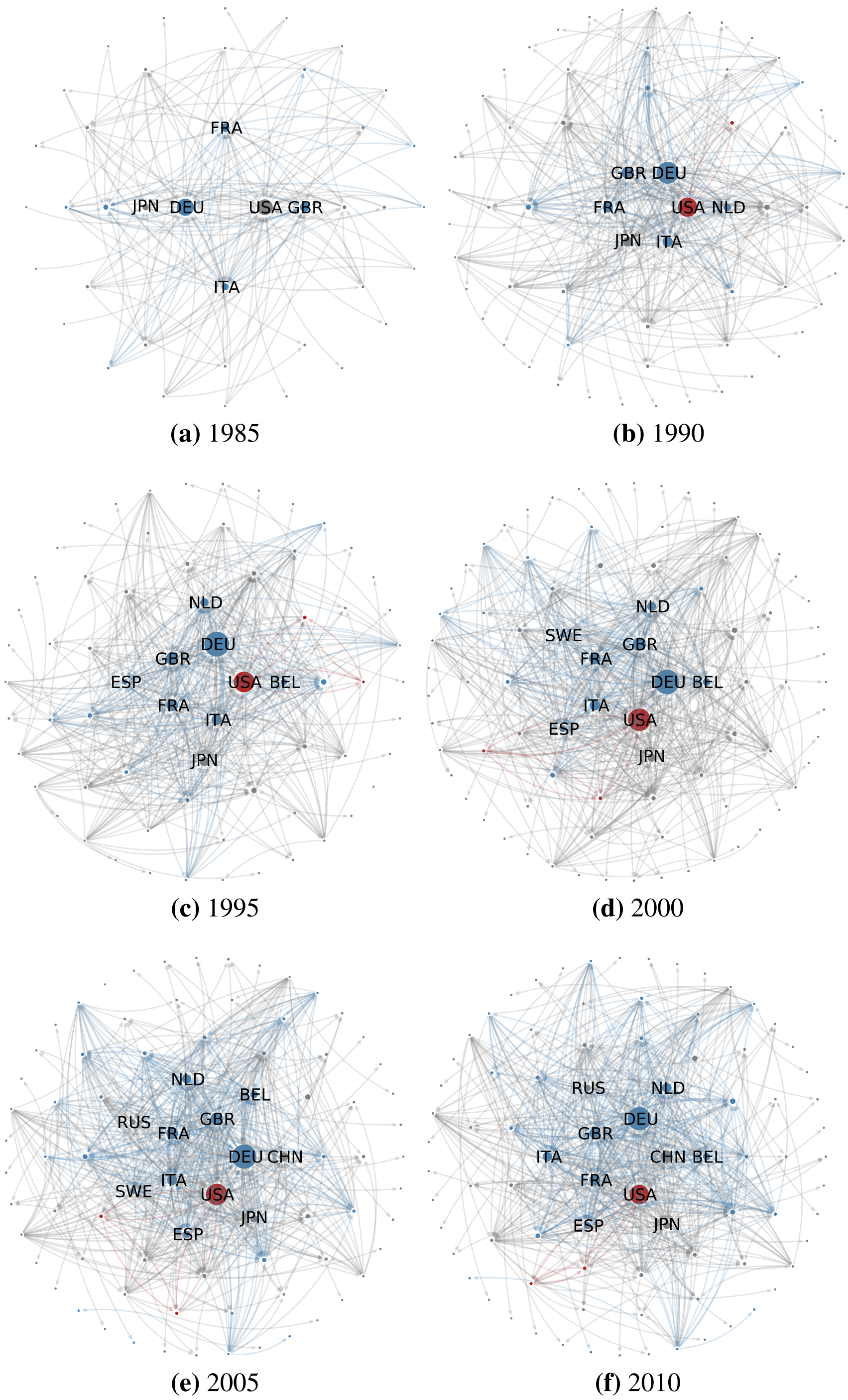
[(a)] The between correlation is defined as the correlation between the means of each country-pair;
[(b)] The within correlation is the correlation between the demeaned values for all country-pairs.

The last two columns further decompose the overall correlation into the correlation between the means for each country (between), and that of the demeaned series (within). It shows that the positive overall correlation is the result of the high correlation of the mean values, while the within-variation is close to zero. In other words, the choice of indicator will matter a lot in time-series or panel data studies that use the variation over time. This could lead to substantial differences in fixed effects studies, that use only the within variation, or for example in the analysis of the effect of institutional integration on actual economic integration (cf. Dorrucci et al., 2004).

4.4 The effect of the EU and Nafta on the level of integration

This section uses a gravity model approach to look at the effect of the expansions of the European Union (EU) and formation of the Canada-US free trade agreement and the North American Free Trade Agreement (Nafta) on the AEI index. Have these trade agreements increased the level of actual economic integration for the participating countries?

Head and Mayer (2013) give an extensive overview of gravity models, their theoretical underpinnings and the econometric issues that arise when estimating them. In order to be consistent with economic theory, the structural gravity model corrects



(a) 1985 **(b)** 1990

(c) 1995 **(d)** 2000

(e) 2005 **(f)** 2010

Figure 4.9: Plot of the AEI network over time.
The nodes represent countries. The higher the weighted indegree, the bigger the node and the closer toward the middle it lies. Nodes are colored blue (red) when the country is a member of the EU (Nafta), as are the edges between members of the EU (Nafta).

Table 4.3: Effect of the EU and Nafta on Actual Economic Integration

Fixed effects	(1) Sy-Ty	(2) Sy-Ty	(3) Sy-Ty	(4) S-T	(5) S-T	(6) S-T
log(distance)	−0.063	−0.057	−0.064	−0.096	−0.090	−0.095
	(0.002)[a]	(0.002)[a]	(0.001)[a]	(0.002)[a]	(0.002)[a]	(0.002)[a]
Contiguity	0.426	0.401	0.414	0.397	0.377	0.384
	(0.006)[a]	(0.006)[a]	(0.006)[a]	(0.006)[a]	(0.007)[a]	(0.006)[a]
log(GDP_s)	-	-	-	−0.087	−0.119	−0.084
				(0.004)[a]	(0.005)[a]	(0.004)[a]
log(GDP_t)	-	-	-	0.006	0.002	0.006
				(0.002)[a]	(0.003)	(0.002)[a]
Colony	0.032	0.036	0.034	0.051	0.056	0.054
	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.004)[a]	(0.003)[a]
Com. col.	0.013	0.006	0.013	−0.004	−0.009	−0.005
	(0.007)[c]	(0.008)	(0.007)[c]	(0.007)	(0.008)	(0.007)
EU	0.124	0.096	-	0.092	0.079	-
	(0.003)[a]	(0.006)[a]		(0.003)[a]	(0.005)[a]	
F_{10} EU	-	0.047	-	-	0.032	-
		(0.005)[a]			(0.004)[a]	
L_{10} EU	-	0.058	-	-	0.052	-
		(0.006)[a]			(0.005)[a]	
EU6	-	-	0.192	-	-	0.165
			(0.007)[a]			(0.007)[a]
EU9	-	-	0.151	-	-	0.114
			(0.006)[a]			(0.006)[a]
EU15	-	-	0.148	-	-	0.130
			(0.005)[a]			(0.004)[a]
EU27	-	-	0.048	-	-	0.036
			(0.005)[a]			(0.004)[a]
NAFTA	0.786	0.306	0.794	0.744	0.287	0.753
	(0.014)[a]	(0.031)[a]	(0.015)[a]	(0.015)[a]	(0.030)[a]	(0.014)[a]
F_{10} NAFTA	-	0.463	-	-	0.438	-
		(0.026)[a]			(0.024)[a]	
L_{10} NAFTA	-	0.253	-	-	0.268	-
		(0.034)[a]			(0.033)[a]	
nObs	180349	117385	180349	128758	88671	128758

Sender-year and target-year fixed effects (1-3) and sender and target fixed effects (4-6) regression of the log of the AEI index on RIA dummies and control variables. Standard errors (between brackets) are corrected for the uncertainty of the AEI index. [a], [b] and [c] denote significance at 1%, 5% and 10% level.

for (time-varying) multilateral resistance terms by using sender-year and target-year fixed effects. Seeing that the number of sender countries is limited to the OECD members, the number of sender-year dummies is only 1012 as opposed to the 6727 target-year combinations. By first demeaning the dependent and explanatory variables in the target-year dimension the number of dummy variables in the model is severely reduced, making the estimation process fairly straightforward. As a robustness check, the model is also computed using the more simple sender and target fixed effects.

It is interesting to note that the problem of the overabundance of zero-flows in trade data does not occur when using the AEI index, since its values are not truncated at zero. As the index values are all strictly greater than zero, they can simply be log-transformed. In addition, because the AEI is corrected for the size of the sender country, the issue of heterogeneity which complicates the regressions that use trade flows also does not come about, invalidating the need for a Poisson-type estimation. Instead, a simple log-linear model corrected for fixed effects suffices.

The standard deviations reported in table 4.3 also take uncertainty of the AEI index into account. When the state-space model is estimated using Gibbs sampling, it returns a set of random draws from the distribution of the AIE index. These draws also correctly reflect the time-dependence in the index. By estimating the gravity model using Gibbs sampling and using a different draw from AEI's distribution in each iteration, the standard deviation of the gravity model's parameters can be corrected for the uncertainty of the AEI index (Standaert, 2014).

Finally, it bears repeating that these estimations results represent the effect of these trade agreements on the overall level of integration, as opposed to the traditional gravity model that focuses solely on bilateral trade.

The control variables come from CEPII's gravity dataset (Head et al., 2010). While they conform to those listed in Head and Mayer (2013), the size of of the coefficients can differ considerably. For example, according to the first column of table 4.3 an increase in the distance of 1% lowers integration by 0.06% which is significantly lower than the 1.1% most other authors find. On the other hand, sharing a border increases integration with more than half $(100 \times (e^{0.426} - 1)\% \approx 53\%)$ which is on par with what is found in the literature. The level of integration between a colonial power and its former colonies is on average 3% higher, but having been colonized by the same country does not significantly affect integration. This last result is not particularly surprising seeing that trade between non-OECD countries is not covered. The GDPs of sender and target are completely captured by the sender-year and target-year dummies, but not when using the more simple sender and target fixed effects. These show that larger sender countries are less open, while large target countries attract higher levels of integration. Nevertheless, the effect of GDP is very small.

To capture the effect of the trade agreements, the *EU* dummy is one when both countries are members of the European Union and the *Nafta* dummy does the same for the Canada-US and North American free trade agreements. The first column (table 4.3) shows that members of the EU are on average 13% more integrated, while the level of integration between members of Nafta is 119% higher. These coefficients fall well within the bounds of what other structural gravity models have found (Head and Mayer, 2013).

To control for the possibility that the agreements were closed between countries that were already highly integrated, F_{10} EU and F_{10} Nafta lead the agreements by 10 years, while L_{10} EU and L_{10} Nafta capture the long term effects of the agreement by lagging by 10 years. In the case of Nafta, they show almost half of the effect (59%) is because the member states were already highly integrated. However, the agreement also succeeded in raising the level of integration both in the short (36%) and long term (29%).

The interpretation in the case of the EU is slightly less straightforward because the agreement predates the period studied. As a result, the leading EU variable (L_{10} EU) only controls for the level of integration of the new EU25-EU27 member states.[9] Similarly, the latest expansions of the EU are too recent, which is why the lagged values (F_{10} EU) only captures the long-term effect of the EU12-EU15 expansions.[10] They show that the EU25-27 enlargement included countries that were already more highly integrated (5%), but not as much as in the case of Nafta (both in an absolute sense as well as relative to the overall effect of the EU). Additionally, similar to Nafta, the long term effect of the EU12-15 enlargements (6%) is about half of the short term effect (10%).

Finally, column three separates the different enlargements of the EU. For example,

[9] The Czech Republic, Estonia, Hungary, Poland, Slovakia, Slovenia, Cyprus, Latvia, Lithuania, Malta, Bulgaria and Romania.

[10] Austria, Finland, Greece, Portugal, Spain and Sweden.

EU9 is one when both countries are member of EU9 but not when both are also members of the original EU6, etc. Interestingly, it shows that the effect of the EU12-15 (16%) is identical to that of enlargement to nine members (including United Kingdom, Ireland and Denmark) even though the latter have been in the EU for a decade longer (on average). The level of integration of the founding members is the largest (21%), but we cannot rule out that this might be because these countries were already highly integrated.

Figure 4.9 summarizes these findings by plotting the network over time. Countries that are members of the EU (CUFTA/Nafta) are colored blue (red) as are the edges between two member countries.

4.5 Extensions

The state-space model estimated in this chapter can be extended in multiple ways. An obvious extension is to include a larger number of countries. As more non-OECD countries are added, the quality and availability of data becomes increasingly problematic. A second extension concerns the type of integration studied and the unit of analysis. The state-space model can be used to study potential economic integration, or political integration. With respect to the latter, a powerful advantage of the state-space model is that it can combine different types of data. For example, the model defined in section 4.2 only combines continuous variables, but through the use of latent variables it can easily be extended to combine dichotomous information or a combination of both. The value of the (observed) dichotomous indicator y depends on the value of the (unobserved) continuous latent variable $y^\star$ which in turn is driven by the to-be-estimated level of regional integration RI:

$$y^\star_{i,t} = C + Z * RI_{i,t} + \varepsilon_{i,t} \tag{4.9}$$

$$y_{i,t} = \begin{cases} 1 & \text{if } y^\star_{i,t} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.10}$$

The ability to combine both different types of data means that qualitative data on integration can be added without having to impose a subjective scaling. This means that different aspects of integration can be viewed in a parallel, rather than a sequential way. For example, currency unions can be viewed separately from customs unions. In this way, the index would prevent a one-track, EU-dominated view of integration. Secondly, it also does away with the linear scaling between the different forms of integration as the information contained in the continuous indicators provides a natural scaling. For example, if closing a free trade agreement goes hand in hand with a significant increase in bilateral flows, the scaling parameters C and Z (equation 4.9) will be significantly higher than if it leaves those flows unperturbed.

4.6 Conclusion

Despite a 'spaghetti bowl' of agreements of different types, not so much is known about regional integration. Frequently, this is linked to the absence of a representative and adequate measure thereof.

Regional integration is a complex and multidimensional process, which is the main reason why a systematic standard index of integration is lacking to this day. Even the most basic of definitions of regional integration encompasses many different aspects, increasing the difficulty of finding appropriate data exponentially. The solutions to these problems often undermine the objectivity of the resulting index: different definitions, data and methodologies lead to different results and rankings.

The state-space model can bring some much needed standardization and objectivity to the problem of measuring regional integration. By using the time structure present in the regional integration indicators, it circumvents the problem of missing observations. Moreover, the model is designed to filter out the measurement noise and is able to deal with data of inferior and dissimilar quality. The Bayesian estimation of the model returns the entire probability distribution of the regional integration indicator, making it possible to say whether the change in the index over time is significant or whether the level of integration significantly differs between countries. Moreover, this uncertainty can be taken into account whenever the index is used in subsequent statistical analyses or when constructing a network.

To illustrate the advantages of using the state-space model, we computed the level of actual economic integration of all current members of the OECD towards the rest of the world. The index is based on indicators of international flows of goods, of services, FDI and other financial flows and migration. Using network analysis tools Germany, the USA and Great Britain are revealed to be the most central countries to the integration network. The network also shows the slow rise of China and Russia throughout the last two decades.

In line with the findings throughout the literature, we find a positive effect of the EU and Nafta/Cusfta on the level of economic integration. Partly this is due to a selection effect: countries that joined the EU (5%) and Nafta (35%) were already more likely to be highly integrated. Nevertheless, both agreements also had a positive short term and long term effect on the level of integration: 10% to 16% for the EU and 36% to 75% for Nafta.

Based on this first application of the state-space approach an obvious extension would be to include institutional characteristics to capture a broader concept of integration, or estimate a separate index of institutional integration. A further challenge is the use of the indicator for analytical purposes in view of its estimated character, non-stationarity and endogeneity, which calls for appropriate techniques such as a Bayesian VAR approach. We intend to consider this in future research.

References

De Lombaerde, P., Dorrucci, E., Genna, G. and Mongelli, F.P. (2008) Quantitative monitoring and comparison of regional integration processes: Steps towards good practice. United Nations University - Comparative Regional Integration Studies.

De Lombaerde, P., Dorrucci, E., Genna, G. and Mongelli, F. P. (2011) Composite indexes and systems of indicators of regional integration. In De Lombaerde, Flôres, Iapadre, and Schulz (editors) *The regional integration manual.* Routledge/Warwick studies in globalization.

Dennis, D.J. and Yusof Z.A. (2003) *Developing indicators of ASEAN integration - a preliminary survey for a roadmap.* REPSF Project 02/001.

Dorrucci, E., Firpo, S., Fratzscher, M. and Mongelli, F.P. (2004) The link between institutional and economic integration: insights for Latin America from the European experience. *Open Economies Review*, 15:239–260.

Durbin, J. and Koopman, S. (2012) *Time series analysis by state space methods, 2nd edition.* Oxford University Press, Oxford.

Feng, Y. and Genna, G. (2003) Regional integration and domestic institutional homogeneity: a comparative analysis of regional integration in the Americas, Pacific Asia and Western Europe. *Review of International Political Economy* 10(2):278–309.

Head, K. and Mayer, T. (2013) Gravity equations: workhorse, toolkit, and cookbook. Centre for Economic Policy Research discussion paper 9322.

Heston, A., Summers, R. and Aten, B. (2012) Penn world table version 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hufbauer, G.C. and Scott, J.J. (1994) *Western hemisphere economic integration.* Institute for International Economics, Washington, DC.

Kim, C.J. and Nelson, C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* MIT Press, Cambridge.

Koop, G., Poirier, D.J. and Tobias, J.L. (2007) *Bayesian econometric methods.* Cambridge University Press, New York.

Lancaster, T. (2004) *Introduction to modern Bayesian econometrics.* Blackwell, Oxford.

Mongelli, F.P., Dorrucci, E. and Agur, I. (2005) What does European institutional integration tell us about trade integration. European Central Bank Occasional Paper Series 40.

Standaert. S. (2014) Divining the level of corruption: a Bayesian state-space approach. *Journal of Comparative Economics.*

UN-ESCWA (2006) *Annual review of developments in globalization and regional integration in the Arab countries.* United Nations, New York.

UNECA (2001) *Annual report on integration in Africa. methodology for calculating indices of economic integration effort in Africa.* UN Economic Commission for Africa, Addis Ababa.

UNECA (2002) *Annual report on integration in Africa.* UN Economic Commission for Africa, Addis Ababa.

UNECA (2004) *Assessing regional integration in Africa.* UN Economic Commission for Africa, Addis Ababa.

Verbeek, M. (2010) *A guide to modern econometrics.* John Wiley & Sons, Chichester.

World trade report, 2011. World Trade Organization.

Appendices

4.A Variance decomposition

Table 4.4: Goodness of fit and variance decomposition of the AEI index

variable		$R^2_{overall}$	$R^2_{between}$	R^2_{within}
Manufactured goods	outflow / total flow	0.890	0.919	0.405
Primary goods	outflow / total flow	0.874	0.910	0.433
Manufactured goods	inflow / total flow	0.852	0.909	0.343
Primary goods	inflow / total flow	0.603	0.650	0.061
Services	outflow / total flow	0.523	0.745	0.006
Manufactured goods	inflow / GDP	0.522	0.661	0.254
Debt securities	outflow / total flow	0.485	0.554	0.004
Services	inflow / total flow	0.477	0.728	0.001
Primary goods	outflow / total flow	0.431	0.587	0.009
Debt securities	inflow / total flow	0.425	0.485	0.021
Manufactured goods	outflow / GDP	0.393	0.574	0.214
Primary goods	outflow / GDP	0.325	0.454	0.211
Equity	inflow / total flow	0.324	0.368	0.003
Services	inflow / GDP	0.290	0.328	0.004
Equity	outflow / total flow	0.285	0.337	0.015
Foreign population	inflow / GDP	0.259	0.368	0.153
Services	outflow / GDP	0.245	0.286	0.006
Foreign population	outflow / pop	0.238	0.336	0.000
Foreign population	inflow / total flow	0.174	0.201	0.001
Foreign workers	inflow / total flow	0.164	0.160	0.002

Foreign workers	inflow / pop	0.133	0.101	0.252
Foreign workers	outflow / total flow	0.117	0.137	0.002
Foreign population	inflow / pop	0.097	0.110	0.009
Debt securities	inflow / GDP	0.083	0.102	0.001
Debt securities	outflow / GDP	0.063	0.067	0.003
Foreign workers	outflow / pop	0.054	0.066	0.008
Equity	inflow / GDP	0.046	0.052	0.003
Seasonal migration	outflow / pop	0.039	0.065	0.000
Equity	outflow / GDP	0.036	0.040	0.001
Seasonal migration	inflow / total flow	0.032	0.015	0.000
Seasonal migration	inflow / pop	0.031	0.015	0.004
Seasonal migration	outflow / total flow	0.031	0.058	0.002
FDI	inflow / GDP	0.006	0.017	0.001
FDI	outflow / GDP	0.006	0.026	0.001
FDI	outflow / total flow	0.001	0.011	$9.27e-7$
FDI	inflow / total flow	0.000	0.001	$1.7e-6$

List of the R^2 of the measurement equation for each indicator of integration, in descending order of their goodness of fit. The last two columns decompose the overall R^2 into its between and within components (Verbeek, 2010).

5 | Historical trade integration - Globalization and the distance puzzle in the long twentieth century[1]

Abstract

In times of ongoing globalization, the notion of geographic neutrality expects the impact of distance on trade to become ever more irrelevant. However, over the last three decades a wide range of studies has found an increase in the importance of distance during the second half of the twentieth century. This chapter tries to reframe this discussion by characterizing the effect of distance over a broader historical point of view. To make maximal use of the available data, we use a state-space model to construct a bilateral index of historical trade integration. Our index doubles to quadruples yearly data availability before 1950, allowing us to expand the period of analysis to 1880-2011. This implies that the importance of distance as a determinant of the changing trade pattern can be analyzed for both globalization waves. In line with O'Rourke (2009) and Jacks et al. (2011), we find that the first wave was marked by a strong, continuing decrease in the effect of distance. Initially, the second globalization wave started out similarly, but from the 1960s onwards the

[1]This chapter is the result of joint work together with Stijn Ronsse and Benjamin Vandermarliere.

importance of distance starts increasing. Nevertheless this change is dwarfed by the strong decrease preceding it.

Keywords: Trade integration; Globalization; Distance puzzle; State-space model.
JEL: F15; C4; F14.

5.1 Introduction

Over the past two centuries, globalization and the increase in international trade in goods and services has dramatically altered living conditions around the world for billions of people. Understanding the intricacies of the changes in the worldwide trade pattern is therefore of key importance. From as early as the 1980s, authors have visualized international trade using the instruments of network science. By representing countries as nodes and capturing their trade relations by drawing a link (or edge) between a pair of countries (or dyads), disaggregated trade data can be amalgamated into a complete overview of the worldwide trade network.

During the last three decades, gravity models have been used to study the impact of distance on the worldwide trade network. The theory of geographic neutrality predicts that as the world becomes more globalized, the effect of distance on the choice of trading partners would become less important. However, for the second half of the twentieth century, a period marked by ever increasing worldwide trade, the opposite pattern emerges: distance is becoming more important (e.g. Leamer and Levinson, 1995; Schiff and Carrere, 2003). The research question subsequently shifted towards the causes of this *distance puzzle*. Proffered explanations range from sample selection issues (Brun et al., 2005), the choice of estimation model (Bosquet and Boulhol, 2013) to the overall methodology used (Disdier and Head, 2008). This would imply that the distance puzzle would also emerge when the same approach is used when studying globalization in the first half of the twentieth century.

An alternative explanation coming from the field of economic history is that the increasing importance of distance is not caused by some aspect of the analysis framework, but is effectively a feature of the worldwide trade pattern. Globalization in the late 19^{th} to early 20^{th} century (the first wave) was driven by decreasing trade costs. Whereas the second wave in the latter half of the 20^{th} century was induced by changes in the productive capabilities of countries (Jacks et al., 2011) or geopolitical changes centered on Western Europe and North America (O'Rourke, 2009). This means that while geographic neutrality would increase during the first globalization wave, such a pattern would not necessarily be present during the second wave.

This chapter contributes to this literature by expanding the analysis of the worldwide trade network to the period 1880-2011, enabling a direct comparison of the distance effect during both globalization waves. In order to do this, we have to overcome the problem that data availability worsens significantly before the 1950s. Both network and gravity models impose high demands in terms of data availability. Constructing a proportionally weighted network requires data on imports and exports for each dyad as well as the GDPs of each country. Incomplete data makes it impossible to tell whether a change in the network is structural or a result of a change in data availability. The gravity models run into similar problems, especially when the Head-Ries index (HRI) of trade integration is used, since it requires either a measure of internal trade, the GDPs of both countries or extensive tariff data (Head and Mayer, 2013). Because this data is more readily available from the 1950s onwards, most studies are limited to this period leaving out the first globalization wave.

To deal with the availability problem, we propose an alternative indicator of trade integration. A state-space model is used to combine several indicators of the level of trade integration into one overall index: the historical trade integration (*hti*) index. Because of the way it handles missing observations, the state-space model uses differences in availability in an offsetting rather than a reductive way. In other words, differences in availability can compensate for each other instead of reducing the dataset to instances when only all data is available. Gaps in one measure can be imputed automatically using information in the others without imposing strict assumptions on or ad hoc manipulations to the data. This allows us to more than double data availability in the period 1880-1914 and extend the analysis to the period 1880 to 2011, covering both globalization waves. Moreover, by combining the correlates of war bilateral trade data with data from the RICardo project and the IMF's direction of trade statistics, the *hti* index also covers a large fraction of colonial trade.

The *hti* index is subsequently used as the dependent variable in a gravity model to study how the importance of distance evolves over time. In line with O'Rourke (2009) and Jacks et al. (2011), we find a small increase in the importance of distance from the 1960s onwards, but show that this is dwarfed by the sharp decrease during the first globalization wave. In other words the behavior of distance during the second globalization wave is not puzzling when considered in its historical context.

The remainder of this article is organized as follows. The next section provides a short overview of the literature on globalization and the distance puzzle. This is followed by a detailed discussion of the construction of the historical trade index and how it compares to other measures of trade integration. The index is then used in a benchmark gravity model after which we look at the effect of distance during both globalization waves.[2]

5.2 Historical framework

Since the 1980s the term globalization has been used in a myriad of scientific disciplines, each using its own definition(s). A common denominator to most definitions is the shift of economic transactions from the local towards the global market. From this point of view, our dataset can be divided into three distinct phases: 1880-1914 and 1945-1995 were marked by increasing globalization, while the Interbellum (1919-1939) was a period of de-globalization.

There has been much discussion on the timing of the first phase of globalization. Some authors believe it started at the end of the 19th century (e.g. Estevadeordal et al., 2002; Dilip, 2003). Others believe that it started from as early as the 1840s

[2] The *hti* index is made available at: `http://www.sherppa.ugent.be/hti/hti.html`.

(e.g. O'Rourke and Williamson, 2004; Jacks et al., 2010). Nevertheless, both recognize that the end of the 19th century was part of the first globalization wave and that the decreasing importance of trade costs, brought on by political and technological improvements, were an important factor at the time (O'Rourke, 2009). With Great Britain in the lead, the mercantilist era was replaced by the idea of a more free trade regime. The European colonizers also imposed this new trading regime on their colonies and even forced independent countries to open up their trade. Technological progress, such as the use of steam engines and the construction of an extensive railway network significantly reduced trading costs. At the same time, the gold standard offered a stable international trading climate (Crafts, 2004).

This liberalizing trend was undone by the first World War and the subsequent conference of Versailles which did little to stabilize international relations. The situation was exacerbated by the Great Depression and the protectionist policies it induced. At the time, the United States took over the leading role in the world economy but failed to further the free trade agenda and could not pull the world economy out of the recession. World War II strengthened the anti-imperialist nationalist and communist states, the disintegrating effect of which lasted till the 1990s. On top of that, globalization was countered by the use of higher tariffs in support of import substitution policies, mostly by newly decolonized countries (Findlay and O'Rourke, 2007). Inspired by Prebisch-Singer motives, the increased tariffs' magnitude was such that it even raised the average world tariff (Lampe and Sharp, 2013). As a result, the post-war efforts to improve international relations, with for example the GATTs and WTO, had a more regional character limited to Western Europe and North America. Intensification of trade relations took place in these regions, but did not extend to the rest of the world (O'Rourke, 2009; Irwin and O'Rourke, 2011). Because of this, O'Rourke concluded that the second wave of globalization, in contrast to the first one, was not driven by a reduction in trade costs, but by geopolitical factors.

5.2.1 Geographic neutrality and the distance puzzle

Geographic neutrality states that the effect of distance on the trade patterns fades as the world becomes more globalized. The pinnacle of this process is the theoretical ideal of a trade pattern that is completely unaffected by distance, in the words of Frances (1997): the death of distance. However, as Leamer and Levinson (1995) discovered, analyses of the second globalization wave found that distance is becoming more important. This distance puzzle was confirmed in many subsequent studies, as illustrated by Disdier and Head (2008) whose paper offered an overview of 1467 distance estimates from over a hundred papers. The robustness of this finding was surprising given that distance serves as a proxy for (ice-berg type) trade costs in gravity models (cf. Head and Mayer, 2013) and those were assumed to decrease over time.

The economic research into the explanations of this apparent contradiction can broadly be categorized into three groups. The first group attributed the distance puzzle to sample selection and the level of aggregation. For example, it was suggested that the distance puzzle only manifested itself in the trade flows of developed countries but not in the case of developing countries (Brun et al., 2005; Boulhol and De Serres, 2010). Alternatively, Larch et al. (2013) and Bosquet and Boulhol (2013) suggested that the heterogeneity of exporters or of trade flows in general was to blame. Arribas et al. (2011) proposed the construction of a specific integration indicator that takes this country-level heterogeneity into account. Similarly, Bleaney and Neaves (2013) posited that access to the sea, remoteness and land area caused a divergence in the effect of distance. Others claimed that the distance effect could be explained by the use of aggregated trade data and suggest the use of data on the sectoral level (e.g. Siliverstovs and Schumacher, 2009; Berthelon and Freund, 2008).

Besides problems with sample selections, the second group assumed that the used estimation technique has distorted the coefficient on distance. Silva and Tenreyro (2006) and Bosquet and Boulhol (2013) questioned the appropriateness of OLSestimators, arguing that the Poisson pseudo-maximum likelihood (PPML) estimator should be used. Additionally, by log-transforming the dependent variable, dyads with zero trade were removed from the dataset (Coe et al., 2007). While the coefficients changed little or even increased slightly when PPML was used, correcting for zero-trade flows caused the coefficient on distance to decrease during the second half of the 20^{th} century.

The third group of arguments blamed the distance puzzle on wider methodological issues. For example, Disdier and Head (2008) argued that the standard gravity model can only measure relative transport costs. Assuming that the effect of globalization is evenly spread among the different trading partners, it is likely that the coefficients on distance and trade costs remain stable over time. Other arguments were offered by Buch et al. (2004), who posited that the effect of globalization was channeled through the constant gravity term. Schiff and Carrere (2003) suggested that instead of overall trade costs the focus should be on the relative evolution of its components. Finally, Lin and Sims (2012) suggested that the distance puzzle could be explained by the difference between the extensive and intensive margins of trade. They reasoned that many of the new long distance trade links were small in volume, while the opposite held for short trade links.

An alternative to these three groups of explanations comes from the field of economic history. Jacks et al. (2011) reasoned that the behavior of distance during the second globalization wave makes sense when looked at from a wider historical perspective. They argue that globalization in the pre-World War period was driven by decreasing trade costs, leading to a decrease in the importance of distance during this period. On the other hand, during the second globalization wave technological progress and economies of scale increased the productive capabilities of countries, causing the increase in worldwide trade. Decreasing trade costs played a much smaller role, explaining the lack of an increase in geographic neutrality. Using wheat prices of several cities in the United States and Europe, Jacks (2009) went on to show the distinct effect both globalization waves had on transport costs. Subsequent research by Jacks et al. (2010) looked at the endogenous nature of trade and trade costs. They found that the reduction of freight rates in the last quarter of the 19th century was caused by the ongoing globalization wave and not the other way around. Nevertheless, when Jacks et al. (2011) used a broader definition of trade costs, they found that trade costs were a determining factor of the first globalization wave.[3] For a comprehensive overview of the empirical analyses of the determinants of trade costs, we refer to the chapter of Lampe and Sharp (2015) in the Handbook of Cliometrics.

In summary, if the explanation of O'Rourke (2009) holds true, an expansion of the gravity model analysis to include both the globalization waves should reveal the different nature of both waves. On the other hand, if the distance puzzle is caused by sample selection, the estimation technique or the overall methodology, it should also be present during the first globalization wave.

5.3 Measuring historical trade integration

The definition of historical trade integration used in this chapter is based on that of Actual Economic Integration by Mongelli et al. (2005, p.6): *"the degree of inter-penetration of economic activity among two or more countries [...] as measured at a given point in time."* The main difference is that because of data limitations, the historical trade integration index only focuses on traded goods.

Throughout this section, the new index will be compared with other measures used in the literature. In decreasing order of availability, these are exports over total exports; exports over GDP of the sender country (e.g. Fagiolo et al., 2008); the sum of exports and imports over GDP of the sender country (e.g. Arribas et al., 2011); and the Head and Ries Index (HRI) of integration (Head and Ries, 2001), which compares the bilateral trade flows with the level of internal trade of both countries.[4]

[3]In addition to freight rates and distance, Jacks et al. (2011) also controls for tariffs, the gold standard, empire membership, railroad infrastructure, exchange rates, common language and shared borders.

[4]$\sqrt{X_{ij}X_{ji}/\left(X_{ii}X_{jj}\right)}$, with X_{ij} the exports from i to j and X_{ii} the internal trade in country i. Internal trade is usually approximated by subtracting exports from GDP, even though this can cause negative values for small open economies. Alternative solutions include using tariff data (Head and Mayer,

5.3.1 Indicators of trade integration

To measure the level of trade integration we collected four indicators that reveal the importance of the bilateral trade flows for the sender country. The trade flows are normalized to correct for differences in scale, since for example the importance of a million dollars worth of imports will be starkly different in the case of Latvia as opposed to the United States. Defining $X_{ij,t}$ as the total exports from the sender i to target country j in year t, these measures are:

$$y_{ij,t} \equiv \{\frac{X_{ij,t}}{\sum_j X_{ij,t}}, \frac{X_{ji,t}}{\sum_j X_{ji,t}}, \frac{X_{ij,t}}{GDP_{i,t}}, \frac{X_{ji,t}}{GDP_{i,t}}\}.$$

Firstly, the level of trade integration is considered high when a significant fraction of total exports go to, or imports come from, a single partner country. This normalization has the advantage that it can be computed using only trade data, but has the weakness that it does not take the overall openness to trade into account. For this reason, the last two indicators normalize import and export flows using the GDP of the sender country. However, because of the additional need for GDP data, the availability of the latter indicators is significantly lower.

To the extent that all four indicators give a similar signal the resulting index will have small confidence intervals.[5] However, when these indicators start to diverge the standard deviation will enlarge, reflecting the underlying uncertainty of the indicators. For example, in the early sixties Russia imported between one and two million dollars from Pakistan, but exported nothing. Using only exports or imports would give a very skewed view of trade relations and using the sum of both misrepresents the ambiguity of the data. In line with the critique of Morgenstern (1962), the data on trade flows and GDP is not treated as an 'observed fact,' but rather as an estimate with a certain (and sometimes severe) measurement error. If different indicators give an opposing signal, the indexation method will treat this informa-

2013).

[5]Since we will estimate this model using Bayesian techniques it would be more correct to use the term *highest posterior density intervals*, but for readability's sake, we will use *confidence interval* throughout this chapter.

tion as unreliable. The added uncertainty of both the underlying data as well as the indexation method itself is then taken into account in subsequent analyses.

The historical import and export data came from three sources: the Correlates of War (COW) bilateral trade database version 3.0 (Barbieri et al., 2009; Barbieri and Keshk, 2012), the Research on International Commerce (RICardo) database[6] and the IMF's Direction of Trade Statistics (DoTS). Historical GDPs were provided by the Maddison project (Bolt and van Zandan, 2013) and supplemented with data from the Penn World Tables 8.0 (Feenstra et al., 2013).

In accordance with the *Real Openness* measure of Alcalá and Ciccone (2004), trade flows were measured in exchange rate converted US dollars while GDP was measured in PPP converted US dollars. As Alcalá and Ciccone (2004, p.4) show, using exchange rate converted GDPs (like Klasing and Milionis, 2014) makes the measure of openness depend on the level of the nontradable good prices. Especially in the case of developing countries, exchange rates conversion will underestimate the GDP (the Balassa-Samuelson effect) and trade shares will be overestimated.

Following de la Escosura (2000, p.4), the trade flows and GDPs were measured in current dollars instead of the constant 1990 US dollars (or Geary Khamis dollars), because only the former can correctly compare any pair of years in the time span. In order to get GDPs in current PPP, we used Klasing and Milionis (2014) method of multiplying the GDPs in Geary Khamis dollars with a GDP deflator.[7] The deflator was provided by Williamson (2015) and population data came from COW's National Material Capabilities database version 4.0 (Singer et al., 1972) and the Penn World Tables. More information on each of the sources and how they were converted can be found in the appendix 5.A.

This data was collected for 225 countries and territories from 1870 up to 2011,

[6] We are grateful to Beatrice Dedinger (`Beatrice.Dedinger@sciencespo.fr`) for providing access to the unpublised RICardo data. It was converted from pounds to US dollars using the historical exchange rate from Williamson (2015).

[7] De la Escosura (2000) starts with current, exchange rate converted, GDPs and uses the *shortcut method* to compute the current, PPP converted, GDPs. Klasing and Milionis (2014) on the other hand start with Maddison's GDPs in constant, PPP converted, 1990 US dollars and transform it using a GDP deflator in current US dollars. They subsequently transform this series into current, exchange rate converted, US dollars using a similar (but inverted) *shortcut method*.

giving us a total of more than 1.8 million observations. Because most trade data was missing during the World Wars, these periods were left out. It should be noted that as a lot of countries (politically speaking) did not exist at the beginning of the dataset, the total possible number of observations for this period is much lower than the more than 6.6 million suggested by the total number of countries.[8]

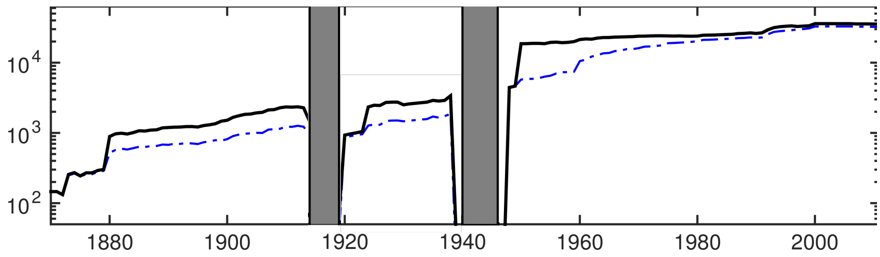
By including the DOTS and RICardo trade data, many colonial countries are covered before their independence. Panel a of figure 5.1 plots the number of observations over time for the entire dataset as well as for the subset of non-colonial countries (the colonial powers and independent states). It shows that the dataset covers colonial trade flows from as early as 1880. Almost half a million of the 1.8 million observations involve a colony and three percent covers trade flows between two colonial countries. While the majority of these trade flows concern the period after World War II, panel b shows that a large number of colonies are covered early on.[9] The large spike in the number of observations in the 1950s is caused by the addition of the DoTS trade data. Finally, it should be mentioned that the trade data only captures the official trade flows between countries. If all trade between two countries passes through a third country (re-exportation) or is smuggled, this will not be captured using this dataset.

5.3.2 The state-space model

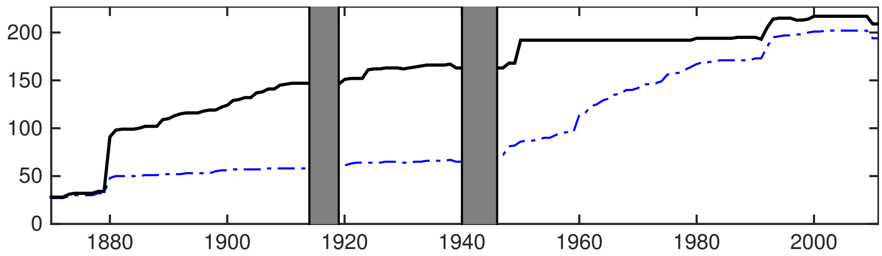
Following the methodology outlined in (Rayp and Standaert, forthcoming), the four indicators were combined into the historical trade integration (*hti*) index using the

[8] n^o of countries × (n^o of countries -1) × n^o of years (excl. World Wars) = $225 \times 224 \times (2011 - 1870 + 1 - 5 - 6)$.

[9] The Overseas Countries and Territories account for the remaining colonies after the year 2000.



(a) Number of observations (logarithmically scaled) per year



(b) Number of countries covered per year

Figure 5.1: Plot of the number of observations and countries in each year for the entire dataset (bold line) and when limited to non-colonial countries (dash-dotted line).

following state-space model:

$$y_{ij,t} = C + Z * hti_{ij,t} + \varepsilon_{ij,t} \quad (5.1)$$

$$hti_{ij,t} = T_t * hti_{ij,t-1} + \nu_{ij,t} \quad (5.2)$$

$$\varepsilon_{ij,t} \sim N(0,H) \quad (5.3)$$

$$\nu_{ij,t} \sim N(0,Q) \quad (5.4)$$

The measurement equation (5.1) states that the four indicators $y_{ij,t}$ try to measure the level of trade integration between sender i and target country j at time t. Unlike for example a simple average the slope Z and intersect C vary over the different indicators of trade integration. Similarly, the variance of the error term ε can differ over all indicators, in contrast to a principal component analysis where this is kept constant. On the other hand, cross-correlation between the error terms of different indicators is ruled out: $E[\varepsilon^{(k)}, \varepsilon^{(m)}] = 0,\ \forall\ k \neq m$.

The state equation (5.2) allows for the trade index to depend on its previous values in the manner of an AR(1) model (an autoregressive model with one lag). This level of dependence (T_t) is assumed to be the same for all dyads. Allowing it to be different for each country couple adds more than a hundred thousand parameters to the model and slows the regression algorithm down to an infeasible degree.[10] By defining the state equation as an AR(1) process, we implicitly restrict T_t to the [-1,1] interval, including the boundary values. In other words, both stationary and non-stationary values of the *hti* index are allowed, but explosive series are not.

Because of their magnitude and duration, the World Wars were likely to have altered trade relations considerably. This, in combination with the lack of information during the wars, is why they are modeled as a structural break. The level of trade integration before and after each World War was assumed to be uncorrelated and the parameter of time dependence can differ over the three periods (equation 5.5). In this way, the estimation of trade integration before World War I is unaffected by whatever changes happened during the Interbellum or after World War II, and vice versa.

$$T_t = \begin{cases} T_1 & \text{if} \quad 1914 > t \\ T_2 & \text{if} \quad 1918 < t < 1940 \\ T_3 & \text{if} \quad 1945 < t \end{cases} \tag{5.5}$$

The issue of incomplete and missing observations is solved by replacing them with information that is entirely uncertain and does not influence the resulting index: $y = 0$, $var(\varepsilon) = \infty$. This allows the model to run uninterruptedly without fundamentally changing the nature of missing data. This, in combination with the time dependency, enables us to increase the number of countries and years for which the index can be calculated without having to impute or otherwise manipulate the data (Kim and Nelson, 1999; Durbin and Koopman, 2012). This matters especially for those observations where there is only partial information, for example when GDP data is missing. Without this solution for missing observations, either the index cannot be computed for those years (reducing the dataset by more than 20%), or the

[10] Initial tests found that the time-dependency is the same for the vast majority of country couples: 94.4% of the time T_{ij} is not significantly different at the 1% level from T_{jl} with $ij \neq jl$.

resulting indicator runs the risk of being distorted. The state-space model on the other hand can still produce an estimate but will adjust the confidence intervals of this estimate to reflect the lack of a complete dataset (cf. infra, figure 5.2).

This model is estimated using a Bayesian Gibbs sampler algorithm, mainly because of the convenience the Gibbs sampling algorithm provides. This algorithm allows us to split up the computation of a complex (posterior) probability into several much simpler conditional probabilities. For example, if the *hti* index values were known, the state and measurement equations become simple linear regressions. Appendix 5.B provides more information on the estimation procedure and an excellent and detailed introduction to Bayesian Gibbs sampling and state-space model can be found in chapters 7 and 8 of Kim and Nelson (1999).

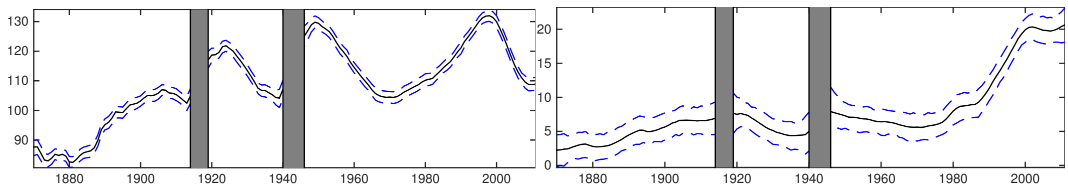
The Gibbs sampler ran for 6000 iterations of which the first 4000 were discarded as burn-in.[11] The remaining iterations were used to reconstruct the posterior distribution of the level of trade integration of each dyad in each year. The index was rescaled to ensure that it has an expected value of zero if there is no trade between the country couple. This was done by generating the expected value of the index when all indicators are zero in each iteration and subtracting it from the *hti* estimate. The resulting index is a continuous variable with values between -0.8 and 189.[12] Following the taxonomy of the levels of measurement, the *hti* index is a ratio variable. While the actual values of the index are meaningless, it can be compared over time and countries without any restrictions. Barring any measurement errors, the index is zero when the underlying indicators are zero. As a result, a doubling of the underlying indicators would result in a doubling of the index value.

[11]The size of the dataset required the use of the resources of the Flemish Supercomputer Center, which was kindly provided by Ghent University, the Flemish Supercomputer Center (VSC), the Hercules Foundation and the Flemish Government – department EWI.

[12]While the underlying indicators are never smaller than zero, the index can still be negative due to the nature of the state-space model, i.e. the fact that indicators are viewed as imprecise measures of the actual level of integration. For example, a positive measurement error on an indicator that is zero would result in a negative *hti* estimate.

5.3.3 The historical trade integration index

By way of illustration, figure 5.2 shows the index values for USA-Mexican bilateral trade from the perspective Mexico (panel a) and the USA (panel b). The figure shows both the expected value of the index as well as its 95% confidence interval. It should be clear from this graph that the level and evolution of trade integration can differ significantly depending on the point of reference. The Mexican-US trade is highly important to the former as its index values lie entirely within the top 1 percentile. From the perspective of the US on the other hand, trade with Mexico only really becomes important from the mid-twentieth century onwards. The divergence in the evolution of the *hti* index values of both countries in the 21^{st} century exemplifies the fact that *hti* measures *relative* trade integration. Trade between the US and Mexico did not decrease after 2000, but trade between China (and to a lesser extent Canada) and Mexico did increase significantly. This led to a drop in the Mexico-US index, but had no effect on the US-Mexico *hti* index values. Furthermore, the widening of the confidence interval immediately after the World Wars illustrates the effect of a decrease in data availability in this period.



(a) Mexico-USA **(b)** USA-Mexico

Figure 5.2: The normalized historical trade index and 95% confidence interval (dotted lines).

The most notable difference between the *hti* index and the other indicators of trade integration is the increase in data availability, especially when compared to the Head-Ries index. When using one of the alternative indicators, overall data availability decreases with 13% in the case of exports over total exports and even 38% ($\approx$ 700,000 observations) when using the HRI. To illustrate, figure 5.3 plots the number

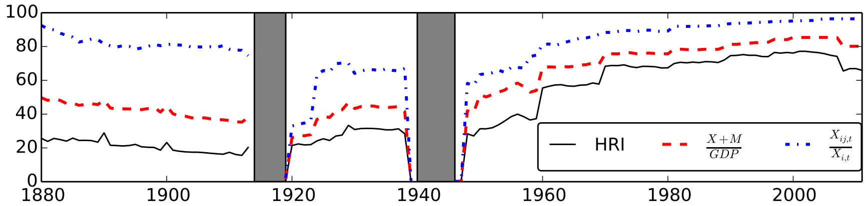


Figure 5.3: Yearly availability of the alternative indicators of integration as a percentage of the availability of the *hti* index.
Plot of the number of dyads covered by the Head-Ries Index (HRI), bilateral openness $((X_{ij,t} + X_{ji,t})/GDP_{i,t})$ and the exports over total exports $(X_{ij,t}/X_{i,t})$ in each year, relative to the number of dyads covered by the historical trade integration index.

of dyads covered by each index over time, expressed as a percentage of the number of dyads covered by the *hti* index. Overall, the increase in data availability when using *hti* grows the further we go back in time. For example, more than half the dyads covered by *hti* are missing when using trade flows over GDP and this rises to 80% when using the HRI.

In order to provide a comprehensive overview of the historical trade integration index we used it to construct a weighted, directed network. Two countries were linked by an edge if their index values were statistically significant, with the index values serving as edge weights (details in appendix 5.C). Figure 5.4 shows the evolution of this network over time. The higher the indegree of a country (the sum of all incoming edges), the more central its position. The larger the pagerank (similar to the indegree, but it gives a higher weight to edges coming from countries that are themselves important (Newman, 2010)), the bigger the size of the country's node. With only a few exceptions, the indegree and pagerank reach the same conclusion on country's centrality. Initially, France and Great Britain were the most central players, but over time the USA superseded both. After World War II, Germany started to overtake both France and Great Britain, rising to the second most central position. The last two decades of the dataset are marked by the rapid ascent of China as one of the most prominent countries in the network.[13] Lastly, as is shown in the appendix 5.4, both globalization waves considerably increased the overall

[13]These and other yearly graphs are made available together with the indicator at `http://www.sherppa.ugent.be/hti/hti.html`.

connectivity of the trade network.

Finally, table 5.1 compares the *hti* index with the indicators of integration it summarizes. The first column shows the R^2 of the measurement equations of each variable, ordered according to their goodness of fit. The second and third column then split this up into the between and within R^2. In addition to this equation-by-equation comparison, the last column compares the contribution of all indicators simultaneously using a dominance analysis based on the R^2 test statistic (Grömpig, 2007). Comparable to the normalized eigenvalues in a principal components analysis, the dominance score shows the relative contribution of each variable to the *hti* index.[14] It comes to the same conclusion as the equation-by-equation R^2. The imports and exports flows divided by total flows are the two most important variables and their between and within R^2 are both high. While the overall match between the flows divided by GDP is much lower, their between match is still acceptable.

Table 5.1: Goodness of fit and variance decomposition of the HTI index

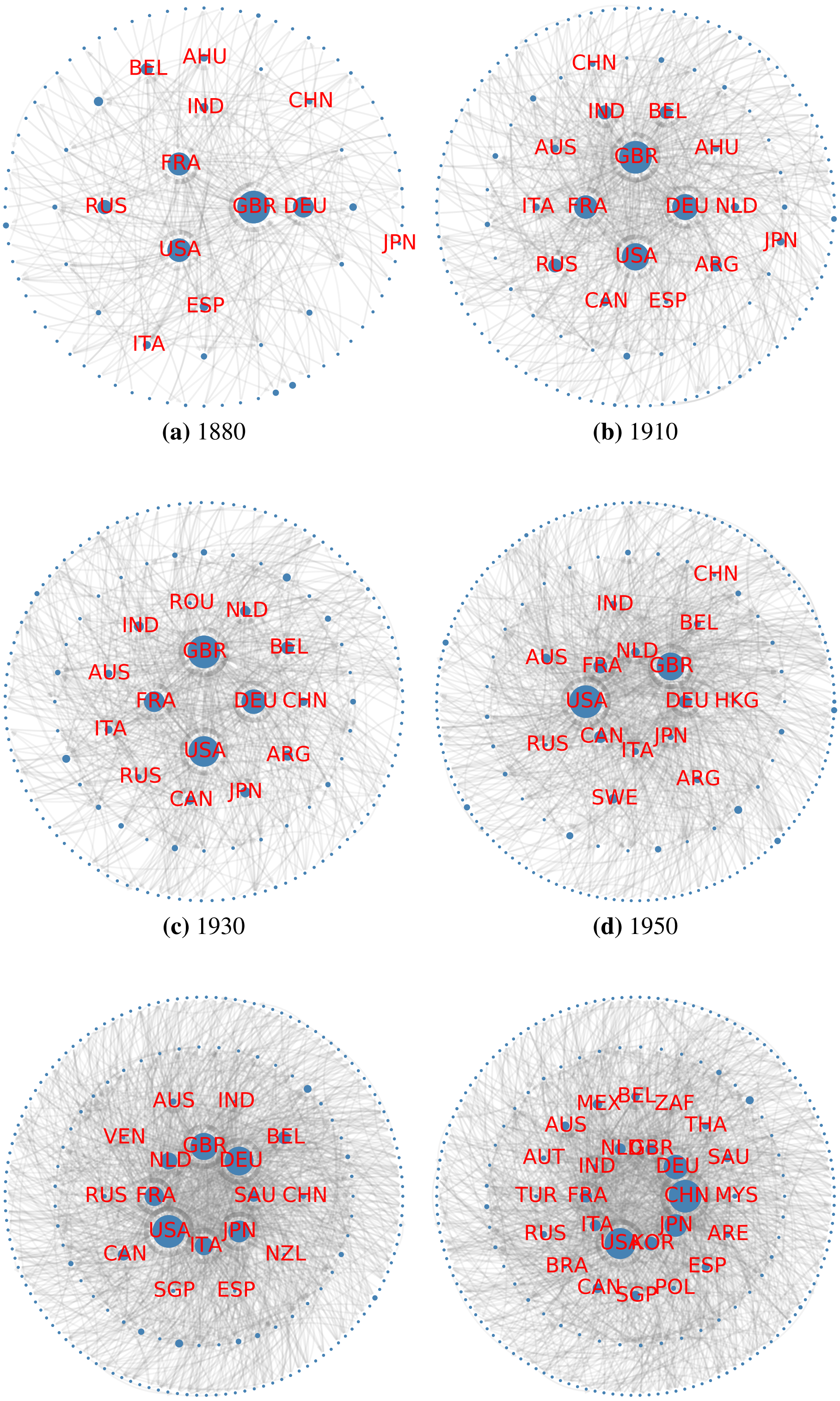
	$R^2_{overall}$	$R^2_{between}$	R^2_{within}	Dominance
Imports / total imports	0.916	0.974	0.799	0.512
Exports / total exports	0.782	0.880	0.602	0.357
Exports / GDP	0.305	0.479	0.131	0.120
Imports / GDP	0.004	0.193	0.002	0.014

List of the R^2 of the measurement equation for each indicator of integration, in descending order of their goodness of fit. The third and fourth columns decompose the overall R^2 into its between and within components (Verbeek, 2010). The last column shows the dominance statistic of each indicator based on the R^2 statistic (Grömpig, 2007).

5.4 Benchmark regressions

To ensure that the *hti* index values conform to expectations and to provide a benchmark for the later analyses, we regressed the log of the *hti* index on a number of

[14]To compute this statistics, the *hti* variable is regressed on all possible $2^k - 1$ combinations of the individual indicators. The dominance score of a variable is the average marginal contribution to the R^2 over all possible models where the variable is included.



(a) 1880 **(b)** 1910

(c) 1930 **(d)** 1950

(e) 1980 **(f)** 2010

Figure 5.4: The historical trade network over time.
The higher the indegree of the node, the closer the node lies to the center. The size of the nodes is determined by their pagerank. The higher the edgeweight, the darker the edge.

economic and political variables using a structural gravity model:

$$\log(hti_{ij,t}) = \alpha \log(\text{distance}_{ij}) + \beta_1 \log(GDP_{i,t}) + \beta_2 \log(GDP_{j,t}) + \beta_3 X_{ij,t} + \mu_{ij,t} + \varepsilon_{ij,t} \quad (5.6)$$

where $\mu_{ij,t}$ is a vector of fixed effects.

In accordance with Baldwin and Taglioni (2006) the GDPs are measured in current USD. $X_{ij,t}$ contains additional control variables, including membership of the European Union (EU) and the North American Free Trade Agreement (Nafta)[15]. f_{10} EU and l_{10} EU are the 10 year leading and lagged variable of the EU membership dummy. This allows a differentiation between the anticipatory/selection, short term and long term effects of signing the trade agreement. The remaining control variables include a dummy capturing the Interbellum and a measure of the completeness of the dataset in each year (hiiAv). The latter is defined as the number of dyads covered in each year, divided by the total possible number of dyads given the number of countries in the dataset in that year. Finally, we included two dummy variables to control for the influence of wars on trade integration: *War* is one when the dyads are engaged in a military conflict, while *Allied* checks whether the countries were on the same side during a conflict. Both variables were constructed using the COW's Inter-State War dataset version 4.0 (Sarkees and Wayman, 2010).

The results are shown in table 5.2. Following Baldwin and Taglioni (2006) and Head and Mayer (2013), the baseline estimates in column 4 and 5 include sender-year and target-year fixed effects to cancel out any time-varying multilateral resistance terms[16]: $\mu_{ij,t} = \mu_{i,t} + \mu_{j,t}$. In order to estimate this many fixed effects, we used a strategy outlined in Guimarães and Portugal (2009) which was adapted to a Bayesian estimation framework (details in appendix 5.D). For completeness sake, column one shows the results using no fixed effect ($\mu_{ij,t} = 0$); column two using sender-target fixed effects ($\mu_{ij,t} = \mu_{ij}$); and column three using sender and target fixed effects ($\mu_{ij,t} = \mu_i + \mu_j$).

A distinct advantage of using the *hti* index as the dependent variable in the gravity model is that its values are not truncated at zero. When using trade flows, almost half of the observations are zero and would be removed when trade is logarithmically transformed. While the estimation procedure can be adjusted to cope with the selection problem, the

[15]Nafta also includes the preceding 1987 Canada-US Free Trade agreement

[16]Multilateral resistance terms are country-specific barriers to trade that in this case are allowed to vary over time.

truncation of the data is harder to solve (Head and Mayer, 2013). In contrast, because it has an interval scale the *hti* index can simply be rescaled to be strictly greater than zero before being log transformed. Furthermore, because the Gibbs sampler returns the entire distribution of the *hti* index for each dyad and year, the gravity model can be adjusted to take into account the uncertainty of the *hti* index estimate (Standaert, 2014).

Table 5.2: Benchmark gravity regression using the *hti* index

			hti		
	(1)	(2)	(3)	(4)	(5)
Distance	−0.026	–	−0.035	−0.035	−0.035
	(0.000)^a		(0.000)^a	(0.000)^a	(0.000)^a
Contiguity	0.124	–	0.120	0.119	0.117
	(0.001)^a		(0.001)^a	(0.001)^a	(0.001)^a
GDP_s	−0.002	−0.010	−0.011	–	–
	(0.000)^a	(0.000)^a	(0.000)^a		
GDP_t	0.026	0.015	0.017	–	–
	(0.000)^a	(0.000)^a	(0.000)^a		
Interbellum	−0.059	−0.006	−0.010	–	–
	(0.002)^a	(0.002)^a	(0.002)^a		
War	−0.020	−0.011	−0.056	−0.038	−0.038
	(0.008)^a	(0.005)^b	(0.007)^a	(0.008)^a	(0.007)^a
Allied	0.094	0.008	0.025	0.026	0.026
	(0.005)^a	(0.003)^a	(0.004)^a	(0.005)^a	(0.005)^a
EU	0.103	0.044	0.064	0.081	0.033
	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.003)^a
f_{10} EU	–	–	–	–	0.012
					(0.002)^a
l_{10} EU	–	–	–	–	0.083
					(0.004)^a
Nafta	0.763	0.135	0.606	0.645	0.052
	(0.012)^a	(0.010)^a	(0.011)^a	(0.011)^a	(0.024)^c
f_{10} Nafta	–	–	–	–	0.507
					(0.017)^a
l_{10} Nafta	–	–	–	–	0.155
					(0.025)^a
hiiAv	−0.382	−0.112	−0.136	–	–
	(0.002)^a	(0.002)^a	(0.003)^a		
Constant	3.486	–	–	–	–
	(0.002)^a				
Fixed Effects	none	sender-target	sender target	sender-year target-year	sender-year target-year
nObs	1,282,178	1,282,221	1,282,178	1,847,771	1,847,771

Linear (column 1) and fixed effects regression (columns 2-5) on the log of the historical trade index with standard errors (between brackets) corrected for the uncertainty of the hti. ^{a}, ^{b} and ^{c} indicate significance at the 1%, 5% and 10% level.

The results are very similar over the different estimation procedures. While the coefficients on the traditional gravity parameters have the expected sign, they are much smaller than when trade flows are used. This is most likely due to a (non-linear) difference in scaling seeing that the relative differences of the parameters remains more or less the same. A decrease in the distance of 1% raises the level of trade integration with 0.04%. Neighboring your trading partner (*contiguity*) further increases this with $100 \times (e^{0.119} - 1)\% \approx 13\%$. Larger partner countries attract more trade, but a rise in the GDP of the home country lowers trade integration. This could be because larger countries tend to be more focused on their internal markets.[17]

Interestingly, both the EU (8%) and Nafta (90%) raised the level of trade between their partner countries. In the case of the EU, the agreement was closed between countries that were slightly more likely to be integrated (1%). The agreement subsequently raised the level of trade integration both in the short term (3.3%) and most importantly long term (8.6%). Nafta on the other hand was closed between countries that were already highly integrated (66%), but it still succeeded in further raising the level of trade integration most of which happened in the long term (17%).

The Interbellum had a negative effect on trade integration, but the effect is rather small (1%). As was expected, being at war significantly lowered integration (-4%), while the opposite held true if the countries were allies in the same war (3%). Finally, we find evidence of selection bias issues in the early values of the *hti* index. The negative parameter on *hiiAv* means that the average index value goes down as the availability of the data increases. The initial values of the *hti* indices (before 1948) are marked by many missing observations, most likely between countries that did not or barely trade with each other. If left uncorrected, this makes the world seem more integrated in the earliest years of our dataset.

5.5 Results

To capture how the distance affects the level of trade we re-estimate the gravity model (6) with the distance variable split into 5 year blocks:

$$\log(hti_{ij,t}) = \sum_{\tau=0}^{n/5-1} \left[\alpha_\tau \log(\text{distance}_{ij}) \mathbb{1}_{\{5\tau < t \leq 5(\tau+1)\}}\right] + \beta X_{ij,t} + \mu_{i,t} + \mu_{j,t} + \varepsilon_{ij,t} \tag{5.7}$$

[17] Since the index is already normalized for the size of the sender country, the GDP of the sender country should actually be left out of the gravity model regressions. However, its inclusion did not significantly affect the results.

α_τ is the distance elasticity of the *hti* values and $\mathbb{1}_{\{5\tau<t\leq 5(\tau+1)\}}$ is an indicator variable separating the (log of the) distance variable into five year blocks. Similar to the regressions in columns four and five of table 5.2, the regression includes fixed effects to control for the time-varying multilateral resistance terms ($\mu_{i,t}$ and $\mu_{j,t}$). Because of the inclusion of sender-year and target-year fixed effects, many of the control variables drop out of the model, most notably the GDP of sender and target.

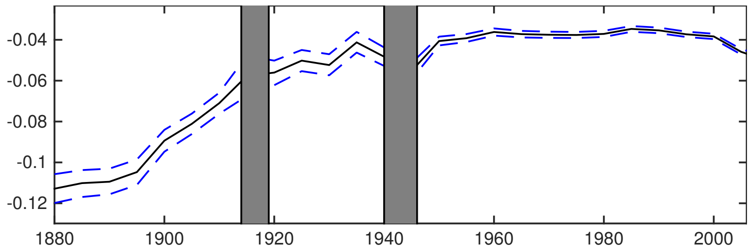
Estimating the distance effect in each year (as opposed to five-year blocks) increased the confidence intervals, but did not change the conclusion. Similarly, the strong increase in data availability during the 1870's significantly distorts the analysis for the first 10 years but leaves the rest unaffected.[18] For the sake of brevity these results are not shown but are available upon request.

Panel a of figure 5.5 plots the distance elasticity (α_τ) over time for the entire dataset. At any time, an increase in the distance will lower the level of trade, explaining the negative coefficients. Overall, the effect of distance on the trade has become smaller over time as the elasticity parameter moves closer to zero. Starting in the 1880s, the importance of distance decreases rapidly, but this process is stopped short by the first World War. While the initial years of the Interbellum still show signs of an increase in geographic neutrality, this trend is reversed in the 1930s. The second globalization wave starts with a very gradual decrease in the importance of distance, but this tapers off by 1960, after which distance slowly becomes more important again.

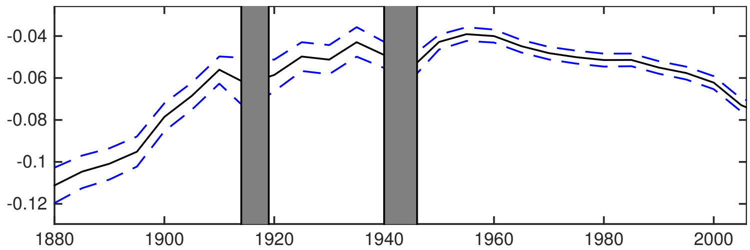
To ensure that the pattern of the distance elasticity is not caused by the increase in the number of countries covered by the dataset, the analysis is repeated while keeping the number of countries constant. To that end, two subsets were defined in which the dataset was limited to countries that persist throughout the entire considered period. One subset starts in 1880 and the other in 1950. Figure 5.6 provides an overview of the countries included. In addition, it also plots the significant trade links in the starting year of the subsets, showing the extent of the dataset in the period with the lowest data availability. The list of countries in each subset is included in appendix 5.E.

The biggest change of keeping the set of countries fixed is that the distance puzzle becomes more pronounced: in both subsets distance becomes more important from the 1960s onwards. This increase in the distance elasticity in the 1880 subset is relative large and ef-

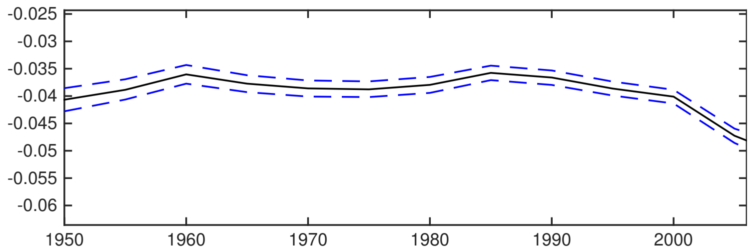
[18] The number of dyads covered increases more than sixfold between 1870 and 1880.



(a) Overall



(b) 1880 subset

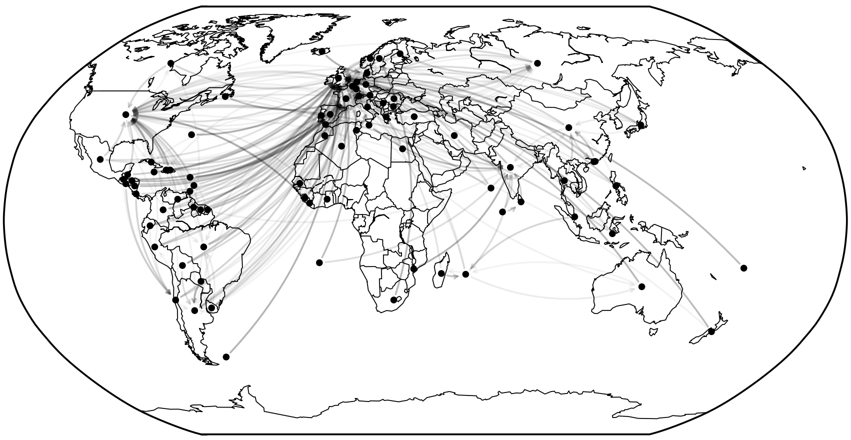


(c) 1950 subset

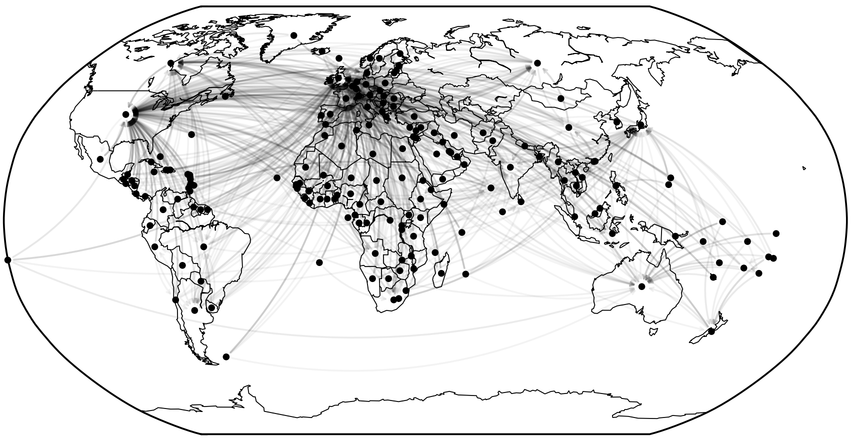
Figure 5.5: The distance elasticity of the historical trade integration index (α_τ) over time

fectively restores the distance elasticity to its pre-World War I levels. However, in the more replete 1950 subset (panel c), the increase is much smaller, mimicking what happens in the overall regression.

These results fall in line with the mechanisms described in O'Rourke (2009) and Jacks et al. (2011), confirming the idea that the behavior of distance should be looked at from a broader historical perspective. The increase in geographic neutrality during the first globalization wave can be explained by the political and technological developments significantly lowering trade costs. The second globalization wave on the other hand was less driven by changing trade costs, but instead by increases in productivity, economies of scale. The geopolitical determinants supporting globalization were centered on Western Europe and North America and counteracted by import-substitution policies in the developing world.



(a) 1880



(b) 1950

Figure 5.6: The coverage of the 1880 and 1950 subsets
Countries included in the subsets are marked by a dot and significant trade links in the starting year by an arrow. The higher the *hti* index the darker the arrow.

5.6 Conclusion

The theory of geographic neutrality predicts that as the world becomes more globalized, the effect of distance on the choice of trading partners would become less important. However, for the second half of the twentieth century, a period marked by ever increasing worldwide trade, the opposite pattern emerges: distance is becoming more important. In this chapter, we contribute to the discussion on the *distance puzzle* by looking at the behavior of distance from a broader historical perspective.

Using an alternative index of trade integration, we expanded the gravity model analysis of the worldwide trade network to the period 1880-2011, enabling a direct comparison of the distance effect during both globalization waves. A number of indicators measuring the importance of bilateral trade were combined into the historical trade index using a Bayesian state-space model. Because of the way it handles missing observations, the state-space model uses differences in availability in an offsetting rather than a reductive way. Gaps in one measure can be imputed automatically using information in the others without imposing strict assumptions on or ad hoc manipulations to the data. This allowed us to more than double data availability in the period before World War I.

Armed with this index, we analyzed the effect of distance on the trade pattern. While we do find that the effect of distance on trade tends to increase from the sixties onwards, this is overshadowed by the strong decrease that precedes it during the first globalization wave. These results support the mechanisms described in O'Rourke (2009) and Jacks (2009) that attribute the distance puzzle to differences in the driving factors of globalization of both waves. During the first globalization wave technological advancements severely reduced the cost of trade, helped along by political developments that broke down many barriers to international trade. The net result of which was a substantial decrease in the importance of distance in choosing trading partners. On the other hand, the second globalization wave was driven by production factors, economies of scale and geopolitical determinants centered on Western Europe and North America. This explains why the decrease in the effect of distance tapers off after 1960. However, we would like to point out that the corroboration of the economic history arguments does not entail that the other hypotheses are without merit. Methodological aspects undoubtably play a role in the explaining part of the distance puzzle.

The index of historical trade integration and plots of the trade network over time are available for download at `http://www.sherppa.ugent.be/hti/hti.html`.

References

Alcalá, F. and Ciccone, A. (2004) Trade and productivity. *The Quarterly Journal of Economics* 119(2):613–646.

Arribas, I., Pérez, F. and Tortosa-Ausina, E. (2011) A new interpretation of the distance puzzle based on geographic neutrality. *Economic Geography* 87(3):335–362.

Barbieri, K. and Keshk, O. (2012) Correlates of war project trade data set codebook, version 3.0.

Barbieri, K., Keshk, O. and Pollins, B. (2009) Trading data: evaluating our assumptions and coding rules. *Conflict Management and Peace Science* 26(4):471–491.

Baldwin, R. and Taglioni D. (2006) Gravity for dummies and dummies for gravity equations. Natinal Bureau for Economic Reserach w12516.

Berthelon, M. and Freund, C. (2008) On the conservation of distance in international trade. *Journal of International Economics* 75(2):310–320.

Bleaney, M. and Neaves, A. S. (2013) Declining distance effects in international trade: Some country-level evidence. *The World Economy* 36(8):1029–1040.

Bolt, J. and van Zanden, J. (2013) The first update of the maddison project; re-estimating growth before 1820. *Maddison Project Working Paper* 4.

Bosquet, C. and Boulhol, H. (2013) What is really puzzling about the 'distance puzzle'. *Review of World Economics* 151(1):1–21.

Boulhol, H. and De Serres, A. (2010) Have developed countries escaped the curse of distance? *Journal of Economic Geography* 10(1):113–139.

Brun, J.F., Carrere, C., Guillaumont, P., and De Melo, J. (2005) Has distance died? Evidence from a panel gravity model. *The World Bank Economic Review* 19(1):99–120.

Buch, C.M., Kleinert, J. and Toubal, F. (2004) The distance puzzle: on the interpretation of the distance coefficient in gravity equations. *Economics Letters* 83(3):293–298.

Carter, C.K. and Kohn, R. (1994) On gibbs sampling for state space models. *Biometrika* 81(3):541–553.

Coe, D. T., Subramanian, A. and Tamirisa, N. T. (2007) The missing globalization puzzle: Evidence of the declining importance of distance. IMF Staff Papers, pages 34–58.

Crafts, N. (2004) Globalisation and economic growth: a historical perspective. *The World Economy* 27(1):45–58.

de la Escosura, L.P. (2000) International comparisons of real product, 1820-1990: an alternative data set. *Explorations in Economic History* 37:1–41.

Dilip, K. (2003) *The economic dimensions of globalization.* Palgrave Macmillan, Hampshire, UK.

Disdier, A.C. and Head, K. (2008) The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics* 90(1):37–48.

Durbin, J. and Koopman, S. (2012) *Time series analysis by state space methods, 2^{nd} edition.* Oxford University Press, Oxford.

Estevadeordal, Frantz, B. and Taylor, A.M. (2002) The rise and fall of world trade, 1870-1939. National Bureau for Economic Research working paper w9318.

Fagiolo, G., Reyes, J. and Schiavo, S. (2008) On the topological properties of the world trade web: A weighted network analysis. *Physica A: Statistical Mechanics and its Applications* 387(15):3868–3873.

Feenstra, R.C., Inklaar, R. and Timmer, M.P. (2013) The next generation of the Penn world table.

Findlay, R. and O'Rourke, K.H. (2007) *Power and plenty: trade, war, and the world economy in the second millennium.* Princeton University Press, Princeton.

Frances, C. (1997) *The death of distance.* Harvard Business School Press, Boston.

Guimarães, P. and Portugal, P. (2009) A simple feasible alternative procedure to estimate models with high-dimensional fixed effects. IZA Discussion paper 3935.

Grömpig Ulrike (2007) Esimators of relative importance in linear regressions based on variance decomposition. *The American Statistician* 61(2):139–147.

Head, K. and Mayer, T. (2013) Gravity equations: workhorse, toolkit, and cookbook. Centre for Economic Policy Research working paper 9322.

Head, K. and Ries, J. (2001) The erosion of colonial trade linkages after independence. *American Economic Review* 91(4):858–876.

Irwin, D.A. and O'Rourke, K.H. (2011) Coping with shocks and shifts: The multilateral trading system in historical perspective. National Bureau for Economic Research working paper w17598.

Jacks, D.S. (2009) On the death of distance and borders: Evidence from the nineteenth century. *Economics Letters* 105(3):230–233.

Jacks, D.S., Meissner, C.M. and Novy, D. (2008) Trade costs, 1870-2000. *The American Economic Review* 529–534.

Jacks, D.S., Meissner, C.M. and Novy, D. (2010) Trade costs in the first wave of globalization. *Explorations in Economic History* 47(2):127–141.

Jacks, D.S., Meissner, C.M. and Novy, D. (2011) Trade booms, trade busts, and trade costs. *Journal of International Economics* 83(2):185–201.

Jacks, D.S. and Pendakur, K. (2010) Global trade and the maritime transport revolution *The Review of Economics and Statistics* 92(4):745–755.

Kim, C.J. and Nelson, C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* MIT Press, Cambrigde.

Klasing, M.J. and Milionis, P. (2014) Quantifying the evolution of world trade, 1870-1949. *Journal of International Economics* 92(1):185–197.

Lampe, M. and Sharp, P. (2013) Tariffs and income: a time series analysis for 24 countries. *Cliometrica* 7(3):207–235.

Lampe, M. and Sharp, P. (2015) Cliometric Approaches to International Trade In Diebolt C. and Haupert, M. (editors) *Handbook of Cliometrics*, Springer, Berlin.

Larch, M., Norbäck, P.J., Sirries, S. and Urban, D. (2013) Heterogeneous firms, globalization and the distance puzzle. IFN Working Paper.

Leamer, E.E., and Levinsohn, J. (1995) International trade theory: the evidence. *Handbook of international economics* 3:1339–1394.

Lin, F. and Sim, N. (2012) Death of distance and the distance puzzle. *Economics Letters* 116(2):225–228.

Mongelli, F.P. Dorrucci, E. and Agur, I (2005) What does European institutional integration tell us about trade integration. European Central Bank Occasional Paper Series 40.

Morgenstern, O (1963) *On the accuracy of economic observations.* Princeton University Press, New Yersey.

Newman, M. (2010) *Networks: an introduction.* Oxford University Press, Oxford.

O'Rourke, K. (2009) Politics and trade: lessons from past globalisations. Bruegel.

O'Rourke, K.H. and Williamson, J.G. (2004) Once more: when did globalisation begin? *European Review of Economic History* 8(1):109–117.

Rayp, G. and Standaert, S. (forthcoming) Measuring actual integration: An outline of a Bayesian state-space approach. In Lombaerde, P.D. and Saucedo, E. (editors) *Indicator-Based Monitoring of Regional Economic Integration*, UNU series on regionalism. Springer, Dordrecht-New York.

Sarkees, M.R. and Wayman, F. (2010) *Resort to War: 1816-2007.* CQ Press, Washington DC.

Schiff, M. and Carrere, C. (2003) On the geography of trade: Distance is alive and well. SSRN 441467.

Siliverstovs, B. and Schumacher, D. (2009) Disaggregated trade flows and the missing 'globalization puzzle'. *Economie internationale* 115(3):141–164.

Silva, J.S. and Tenreyro, S. (2006) The log of gravity. *The Review of Economics and statistics* 88(4):641–658.

Singer, J.D., Bremer, S. and Stuckey, J. (1972) Capability distribution, uncertainty, and major power war, 1820-1965. In Russet, B. (editor) *Peace, war, and numbers*, pages 19–48. Sage, Beverly Hills.

Standaert, S. (2014) Divining the level of corruption: a Bayesian state-space approach. *Journal of Comparative Economics.*

Verbeek, M. (2010) *A guide to modern econometrics.* John Wiley & Sons, Chichester.

Williamson, S. H. (2015) What was the U.S. GDP then? Measuring Worth.

Appendices

5.A Data sources and transformations

Table 5.3: Data sources and transformations

Trade flows - current USD, exchange rate

Source	Original units	Transformations
DoTS	current USD, exchange rate	
COW v3.0	current USD, exchange rate	
RICardo	current British pounds, exchange rate	× pound-dollar rate$^{(a)}$

GDP - current USD, PPP

Source	Original units	Transformations
Maddison	per capita, constant 1990 USD, PPP	× population$^{(b,c)}$
		× 1990 US GDP deflator$^{(a)}$
PWT8.0$^{(b)}$: rGDP^{o}	constant 2009 USD, PPP	× 2009 US GDP deflator$^{(a)}$

$^{(a)}$Williamson (2015)

$^{(b)}$Penn World Tables 8.0

$^{(c)}$COW National Material Capabilities v4.0

5.B Estimating the state-space model

To estimate the state-space model we need to solve for the structural parameters of the state-space model (C, Z, T_t, H and Q) as well as the level of trade integration (hti). While it is possible to maximize the combined distribution numerically for small datasets, using a Gibbs sampler simplifies the estimation procedure considerably by splitting up the process into conditional probabilities.

For example, say we have to draw from the joint probability of two variables $p(A,B)$, when only the conditional probability of $p(A|B)$ and $p(B|A)$ are known. Starting from a (random) value $B^{(0)}$, the Gibbs sampler will draw a first value of A conditional on $B^{(0)}$: $A^{(1)} \sim p(A|B^{(0)})$. Conditional on this last draw, a value of B is drawn ($B^{(1)} \sim p(B|A^{(1)})$) which is in turn used to draw a new value for A ($A^{(2)} \sim p(A|B^{(1)})$. This process is repeated thousands of times, until the draws from the conditional distributions have converged to those of the combined distribution $p(A,B)$. After discarding the unconverged draws (the burn-in), the remaining draws of A and B can be used to reconstitute their respective (unconditional) distributions. Because we are using a Bayesian analysis framework, we have to be explicit about the prior distribution of the parameters. In other words, we have to state what we know about their distribution before looking at the data. Because there is no prior information, we imposed flat priors on Z, C and $log(H)$, meaning that all values in the real space (or real positive space for the variance H) are equally probable.

In the case of the state-space model, the Gibbs sampler consists of two main blocks (Kim and Nelson, 1999):

1. If the level of trade integration (*hti*) were known, the parameters of the measurement and state equations (equation 5.1 and 5.2) could be obtained using simple linear regressions. To ensure the model is identified, the variance of the error term of the state equation (Q) is typically set to 1. Taking for example the situation were there is only one dyad to simplify notation: $hti = (hti_1, ..., hti_n)'$

$$p(T|hti) \propto .5 * \mathbb{1}_{|T|\leq 1} * N(b_T, v_T) \tag{5.8}$$

$$p(Z^k, C^k|hti, y, H) \propto N(b^k_{Z,C}, v^k_{Z,C}) \tag{5.9}$$

$$p(H_{(k,k)}|hti, y) \propto iWish[e^{k\prime}e^k;\, n] \tag{5.10}$$

where *iWish* is the inverse Wishart distribution and

$$\begin{aligned} v_T &= (T'_{t-1}T_{t-1})^{-1} && (5.11)\\ b_T &= v_T * T'_{t-1}T_t && (5.12)\\ v^k_{Z,C} &= (hti'hti)^{-1} * H_{(k,k)} && (5.13)\\ b^k_{Z,C} &= (hti'hti)^{-1} * hti'y^k && (5.14)\\ e^k &= y^k - C^k - Z^k * hti && (5.15)\end{aligned}$$

2. Conditional on the parameters of the state and measurement equations, the distribution of *hti* can be computed and drawn using the Carter and Kohn (1994) simulation smoother.

- *The Kalman filter:* computes the distribution of *hti* conditional on the information in all previous years. Starting from a wild guess, $p(hti_0) = N(0, \infty)$, the following equations are iteratively solved for $t = 1$ to $t = n$:

$$\begin{aligned} a_{t|t} &= E(hti_t|y_1, ..., y_t)\\ &= T * a_{t-1|t-1} + \kappa(y_t - C - ZTa_{t-1|t-1}) && (5.16)\\ p_{t|t} &= V(hti_t|y_1, ..., y_t)\\ &= p_{t|t-1} + \kappa Z p_{t-1|t-1} && (5.17)\end{aligned}$$

  with $\kappa = p_{t|t-1}Z'(Zp_{t|t-1}Z' + H)^{-1}$; and $p_{t|t-1} = Tp_{t-1|t-1}T' + Q$.

- *Simulation smoother:* Draws from the distribution of *hti* conditional on all information in the data and the previous draws. Starting from the last iteration of the Kalman filter, draw $\hat{hti}_n$ from $N(a_{n|n};\ p_{n|n})$ and iterate backwards from $t = n - 1$ to $t = 1$:

$$\begin{aligned} a_{t|n} &= E(hti_t|y_1, ...y_n) \\ &= a_{t|t} + \varsigma(\hat{hti}_{t+1} - Ta_{t|t}) \qquad (5.18) \\ p_{t|n} &= V(hti_t|y_1, ...y_n) \\ &= p_{t|t} + \varsigma(p_{t+1|n} - Tp_{t|t}T' - Q)\varsigma' \qquad (5.19) \end{aligned}$$

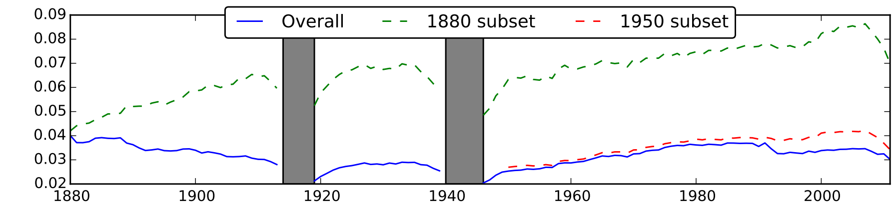
with $\varsigma = p_{t|t}T'p^{-1}_{t+1|t}$; and $\hat{hti}_{t+1}$ a random draw from $N(a_{t+1|n};\ p_{t+1|n})$.

5.C The historical trade network

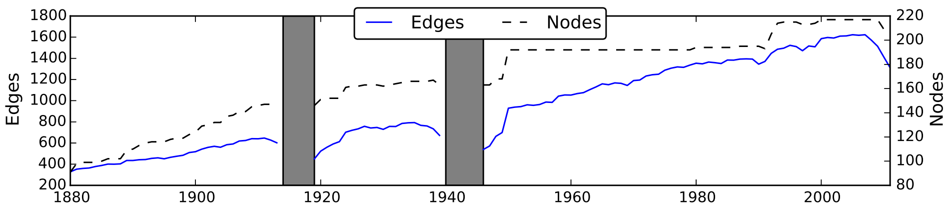
In order to combine the historical trade integration indices into a network, the index values corresponding to countries that are integrated need to be separated from those corresponding to countries that are not. A natural way of making this distinction is to contrast countries that trade with each other ($X_{ij,t} > 0$) to those that do not ($X_{ij,t} = 0$). The problem is that this approach is skewed by a large number of very small non-zero trade flows.

Rather than choosing an arbitrary cut-off value, the *hti* allows us to use significant differences to determine which countries are linked. To start, we used the estimates of the structural parameters of the state-space model to generate index values for a fictional dyad where trade was zero for the entire period. Labeling these observations as $hti_{0,t}$, we defined significant levels of trade in the following way: An edge e from country i to country j exists if, and only if, its level of trade in year t is significantly higher than that of $hti_{0,t}$: $e_{ij,t} = 1 \iff hti_{0,t} < hti_{ij,t}$ in at least 99% of all iterations of the (converged) Gibbs sampler. Using the $hti_{0,t}$ definition, 115,911 edges were identified (6.3% of observations).

Panel a of figure 5.7 shows the overall network density (the fraction of dyads that are connected) gradually decreasing throughout the first globalization wave. In contrast, the trade network becomes increasingly connected during the second globalization wave. As can be seen in panel b, the number of trade links (edges) more or less continuously grows over the entire time-period and is initially offset by the rapid rise in the number of countries. This is especially noticeable when the Soviet Block breaks up in the 1990s, causing a rapid downward shift in the network density.



(a) Density



(b) Number of nodes and edges

Figure 5.7: Network density (panel a) and the number of nodes and edges (panel b) over time.

Similar to the distance regressions, the density was also computed when the number of countries was kept constant using the 1880 and 1950 subsets. This reveals that the decrease in density during the first globalization wave was driven by the addition of new countries. When this is kept constant, the network density almost doubles during the first wave. In addition, it reinforces the effects of the 1930 and 2008 economic crises, both causing a substantial drop in the density. To ensure that these results were not driven by the inclusion of the colonial trade data, the density was also computed using only the official countries according to the COW state system dataset. However, this did not significantly alter the conclusion (available upon request). In other words, once the density is corrected for the increasing number of countries, it conforms to the globalization pattern found in the literature.

5.D Estimating models with high-dimensional fixed effects

Following Guimarães and Portugal (2009), the number of fixed effects can be reduced by half by first demeaning both dependent and explanatory variables in the sender-year dimension, leaving only the sender-target dummies. Using conditional probabilities, the fixed effects (c_i) can be separated from the explanatory variables ($X_{i,t}$), which significantly reduces the size of the matrix that needs to be inverted.

$$y_{i,t} = c_i + X_{i,t}\beta + \varepsilon_{i,t} \quad \text{with } \varepsilon_{i,t} \sim N(0, \sigma^2) \tag{5.20}$$

Equation 5.20 can be estimated using a three-step Gibbs sampling procedure. For example, when using flat (uninformative) priors, the conditional probabilities are:

1. $\beta | c_i, \sigma^2 \sim N(e_\beta, v_\beta)$
   $e_\beta = (X'X)^{-1}(X'(y-c))$ with $\{X\}_{i,t} = X_{i,t}$ and $\{y-c\}_{i,t} = y_{i,t} - c_i$
   $v_\beta = \sigma^2 (X'X)^{-1}$

2. $c_i | beta, \sigma^2 \sim N(\bar{c}_i, \sigma^2/n)$
   $\bar{c}_i = \sum_t^n (y_{i,t} - X_{i,t}\beta)/n$ with n the number of observations of country i

3. $\sigma^2 | beta, c_i \sim \text{iWishart}(e'e, N)$
   $e = y_{i,t} - c_i - X_{i,t}\beta$

5.E Country subsets

Table 5.4: Country Subsets

Group 1: included in 1880< and 1950<				
Algeria	Cuba	Guatemala	Malta	Singapore
Argentina	Denmark	Guyana	Mauritius	South Africa
Ascension	Dominican Rep.	Haiti	Mexico	Spain
Australia	Dutch Antilles	Honduras	Morocco	Sri Lanka
Austria	Ecuador	Hong Kong	Mozambique	St. Pierre and Miquelon
Barbados	Egypt	Iceland	Netherlands	Suriname
Belgium	El Salvador	India	New Zealand	Sweden
Belize	Falkland Isl.	Indonesia	Nicaragua	Switzerland
Bermuda	Fiji	Iran	Norway	Thailand
Bolivia	Finland	Italy	Paraguay	Trinidad and Tobago
Brazil	France	Jamaica	Peru	Tunisia
Bulgaria	French Guiana	Japan	Philippines	Turkey
Canada	Germany	Liberia	Portugal	United Kingdom
Chile	Ghana	Luxembourg	Romania	United States
China	Gibraltar	Macau	Russia	Uruguay
Colombia	Greece	Madagascar	Senegal	Venezuela
Costa Rica	Guadeloupe	Maldives	Sierra Leone	Yugoslavia
Group 2: included in 1950<				
Afghanistan	Congo, Rep.	Hungary	Namibia	Seychelles
Albania	Congo, Dem. Rep.	Iraq	Nauru	Solomon Islands
American Samoa	Cote d'Ivoire	Ireland	Nepal	Somalia
Angola	Cyprus	Israel	New Caledonia	South Korea
Antigua and Barbuda	Czechoslovakia	Jordan	Niger	St. Kitts and Nevis
Bahamas	Djibouti	Kenya	Nigeria	Sudan
Bahrain	Dominica	Kiribati	North Korea	Swaziland
Bangladesh	Equatorial Guinea	Kuwait	Oman	Syria
Benin	Eritrea	Laos	Pakistan	Tanzania
Bosnia	Estonia	Latvia	Palestine	Togo
Botswana	Ethiopia	Lebanon	Panama	Tonga
Brunei	Faroe Islands	Lesotho	Papua New Guinea	Tuvalu
Burkina Faso	French Polynesia	Libya	Poland	UAE
Burma	Gabon	Lithuania	Qatar	Uganda
Burundi	Gambia	Malawi	Rwanda	Vanuatu
Cambodia	Greenland	Malaysia	Saint Lucia	Vietnam
Cameroon	Grenada	Mali	Saint Vincent	Wallis and Futuna
Cape Verde	Guam	Mauritania	Samoa	Yemen
Central African Rep.	Guinea	Mongolia	Sao Tome and Principe	Zambia
Chad	Guinea-Bissau	N. Mariana Isl.	Saudi Arabia	Zimbabwe

6 | Trade integration and trade agreements - resolving the endogeneity problem through a qualitative VAR[1]

Abstract

While the endogeneity of trade and regional integration agreements was established early on, this issue has only been addressed explicitly in gravity models during the last decade and a half. Initial attempts using instrumental variables proved unreliable, causing authors to look for alternative solutions. This chapter brings together the literature on both gravity equations explaining trade and probit regressions explaining the probability of an integration agreement. This is done by estimating them simultaneously in a qualitative vector autoregression model. The qualitative VAR allows us to estimate their interdependence without having to resort to instrumental variables. In addition, the endogenous nature of other control variables like the GDP or the capital labor ratio can be taken into account. Our preliminary findings confirm that an increase in trade raises the probability of an agreement and vice versa, although the response can differ over specific continents. We find a relatively small average treatment effect of RIAs: trade increases with 10% after one year and 40% after five years whereafter it slowly rises to 80% after 35 years.

Keywords: Endogenous trade agreements; Gravity equation; Qualitative choice models;

[1]This chapter is the result of joint work together with Prof. Dr. Glenn Rayp.

Qualitative VAR.

JEL: C11; C25; F14; F15.

6.1 Introduction

Not long after Tinbergen (1962) introduced gravity models to study international trade flows, dummies were added to control for, and measure the effects of regional integration agreements (RIAs).[2] However, the results from these studies have not been very encouraging: depending on the methodology used, the sign and significance of the coefficients on the RIA dummies could change by a wide margin.[3]

The gravity model has evolved strongly since the sixties as its theoretical underpinnings were secured. Starting from a 'naive' log-linearized gravity model, the structural model has been adjusted to take multilateral resistance terms, zero-trade flows and heteroskedasticity into account. At the same time, it became clear that trade and trade agreements are highly endogenous: trading blocs are likely to form along the lines of natural trading partners, i.e. countries that already trade intensively (Krugman, 1995).

In contrast to the large literature on the effects of trade agreements on trade, the literature studying the endogeneity of both has remained limited. Initially, Baier and Bergstrand (2002) and Magee (2003) used an instrumental variables approach, proving the existence of the endogenous relationship. However, estimates of the effect of trade agreements remained unstable and if anything argued against using instrumental variables in cross-sectional studies (Magee, 2003). Baier and Bergstrand (2007) proposed using panel data with either country-year fixed effects or first differences to cope with endogeneity problems. Alternatively, Baier and Bergstrand (2009) used non-parametric matching econometrics to find the right counterfactual to countries that had signed an agreement. Finally, Egger et al. (2011) returned to instrumental variables in a cross-sectional setting. Using a two-part Poisson pseudo-maximum likelihood estimator they controlled for general equilibrium effects and zero-trade flows in addition to the endogenous nature of trade agreements. Overall, the distortion in the effect of RIAs on trade caused by ignoring the endogeneity has been found

[2]Throughout this chapter we will use the term regional integration agreements as a container term for inter- and intra-regional free trade agreements, customs unions, common markets and economic unions.

[3]See Frankel (1997) for an overview of the earlier literature.

to be highly significant, ranging from a 75% increase (Egger et al., 2011) to a quintupling (Baier and Bergstrand, 2007).

An alternative approach to deal with endogeneity could be to use a natural experiment, i.e. a completely exogenous event that led some countries to join while leaving others unaffected. By studying the changes in trade following this event, the effect of this RIA can be analyzed. The problem is that even if such an event can be found, it is not easy to argue that its results can be generalized as the average treatment effect of RIAs worldwide. A vector autoregression model (VAR) on the other hand would allows us to treat both trade and trade agreements as endogenous without having to identify instrumental variables. Instead, the focus lies on the dynamic behavior of both variables which is used to identify their long-term interaction. The only problem is that a VAR model requires continuous variables.

The solution is proffered in the macro-economic literature, where Dueker (2005a) explains how a binary indicator of recessions can be added to a VAR model of the (US) economy. To estimate such a *qualitative* VAR, the indicator variable is first defined in terms of a latent equivalent. In this case, the dummy trade agreements variable is said to depend on the *willingness to sign a trade agreement*. This continuous latent variable can be modeled as endogenous with trade using a normal VAR model. The long term relationship between the variables identified in the VAR can subsequently be used to generate counterfactuals, allowing us to determine the treatment effects of signing a RIA. While the model we present initially ignores zero-trade flows, we show that it can be expanded to deal with both problems simultaneously.

To our knowledge, this is the first time a qualitative VAR has been used to analyze the effect of trade agreements. However, it should be stressed that this chapter is intended as an outline of how the qualitative VAR methodology can be used to study trade and trade agreements, rather than a fully worked out analysis. Instead, our aim is to explain the qualitative VAR, show its place in the trade literature and argue that the model produces sensible results. As Baier and Bergstrand (2009, p. 64) note, there is no well-accepted methodology to asses the impact of trade agreements on trade. Rather than a replacement of the current methodology, the qual VAR should be seen as a way to determine the robustness of earlier findings, specifically the average treatment effect of trade agreements on trade.

The next section continues with an overview of the literature on endogenous trade agreements, after which we discuss the qual VAR methodology. Section 6.5 surveys the results and computes the average treatment effects of a trade agreement. This is followed by a discussion of possible extensions to the model and a preliminary conclusion.

6.2 On the endogeneity of trade and trade agreements

Baier and Bergstrand (2002) and Magee (2003) were the first to the explicitly take the endogeneity of trade agreements into account. The former focused on economic determinants while the latter stressed the importance of political factors.

Using a review of the literature on trade and trade agreements Magee (2003) identified instruments for both. These were then used in two separate IV-regressions explaining either trade agreements or trade. To asses the determinants of trade he used as instruments for trade agreements: 1) the difference in log GDP, 2) the amount of intra-industry trade, 3) the bilateral trade surplus, 4) difference in capital labor ratios and 5) the level of democracy. The number of airports, manageable waterways and wether a country is landlocked were used to instrument trade. The instrumented probit regression explaining trade agreement formation confirmed the natural trading partners hypothesis, i.e. that trade agreements were more likely to form between countries that traded intensively. On the other hand, the instrumented gravity model found a highly volatile coefficient on trade agreements. Depending on the control variables, RIAs were even found to have a significant negative effect on trade. Baier and Bergstrand (2002) on the other hand based their analysis on a general equilibrium model explaining the economic determinants of trade agreement.[4] They warned that (in a cross-sectional framework) allowing a simultaneous effect of trade on RIAs and RIAs on trade (cf. Magee, 2003) resulted in a logical inconsistency; one of the two has to be zero for the probability of having an agreement and the probability of not having an agreement to sum up to one –a necessary condition for a probability. While this ruled out the simultaneity as the cause of endogeneity, other factors (for example including a trade-imbalances variable cf. Magee, 2003) could still cause endogeneity. Using the instrumented RIA variable, the agreements' effect on trade quadrupled. However, further research showed that IV regressions of the treatment effect of RIAs were highly unstable (ranging from -92% to +1100%) and that the instruments's exogeneity was often rejected (Baier and Bergstrand,

[4] Their regressions also included a number of political variables taken from the literature.

2004b).

In response, Baier and Bergstrand (2007) turned to panel data, using fixed effects and first differencing to control for endogeneity caused by selection bias, measurement errors in the explanatory variables and missing variable bias. To be consistent with trade theory, the estimation of the gravity model required country-time fixed effects to control for time-varying multilateral resistance terms[5] in addition to the endogeneity issues mentioned. Furthermore, they argued that using first differences also controls for simultaneity since the natural trading partner hypothesis captures a long term relationship and does not extend to variations in the level of trade. Similar to their findings in the 2002 paper, signing a RIAs caused trade flows to double.

To test the robustness of earlier results, Baier and Bergstrand (2009) turned to non-parametric matching to estimate the average treatment effect of RIAs in cross-sectional data. By matching country-couples with a RIA with a credible counterfactual without one, the effect of an agreement could be computed regardless of self-selection issues or non-linearities. In contrast with the first differences approach, this enabled a computation of the long run treatment effects. In line with their earlier papers, RIAs were found to have doubled trade flows on average.

The latest attempt to model the endogeneity of integration agreements and trade flows explicitly was made by Egger et al. (2011), who returned to an instrumental variable approach using cross-sectional data. Their estimations combined the endogeneity literature with general equilibrium effects of trade agreements and a non-log-linear gravity equation that takes zero-trade flows into account. As instruments for trade agreements they used three dummy indicators indicating: 1) whether one of the countries used to be colony of the other; 2) whether they have a common colonial history; and 3) whether the countries-pair used to be one country. Their structural gravity model was estimated using a two-part Poisson pseudo maximum likelihood estimator with an instrumented RIA variable. The average treatment effect of trade agreements was subsequently computed by using the estimated parameters to generate a counterfactual trade flow. They found that ignoring endogenous selection biased the effect of RIAs downwards by as much 188%. Their average treatment effect of 235% was more than twice as large as was identified in Baier and Bergstrand (2002, 2007, 2009), but it concealed large differences between country couples.

[5] See also Baldwin and Taglioni (2006) and Head and Mayer (2013).

6.3 The Qualitative VAR model

The foremost advantage of using a (qualitative) VAR is that it allows us to treat trade and trade agreements as completely endogenous. In contrast to Baier and Bergstrand (2002, 2004b), Magee (2003) or Egger et al. (2011) there is no need to look for instruments that explain trade while having no effect on trade agreements, or vice versa. Finding instruments for trade or RIAs is difficult as it is hard to rule out that they have no effect on the other variable and any that are found are unlikely to explain a large part of the variation in either variable. Accordingly, Baier and Bergstrand (2007) found that the IV approach produced too unstable estimates of the size of the effect of trade agreements on trade.

In addition, the qualitative VAR model allows us to take the endogeneity of other variables into account. For example, GDP and capital-labor ratios have been shown to affect both trade and integration agreements and are unlikely to remain unaffected by either. For this reason, the qualitative VAR is more appropriate than the multivariate probit since the latter *"is set up to emphasize cross-sectional correlations among a set of qualitative variables and the coefficients on exogenous covariates. VARs, in contrast, are better suited to a small system of endogenous variables and a relatively large number of autoregressive lags"* (Dueker, 2005a, p.97). The VAR allows us to model the endogeneity as autoregressive variables as opposed to autoregressive errors.

Finally, by modeling the interaction between trade and trade agreements dynamically, the logical inconsistency identified in Baier and Bergstrand (2002) can be avoided. Both trade and the willingness to form trade agreements depend on what happened in the past. By definition, trade agreements have a unit root: unless some action is taken by both governments, the existence of a trade agreement today will be the same as that of yesterday. Similarly, shocks to the aggregate trade flows show a high degree of persistence even if particular categories within those flows are more volatile. By modeling their interaction dynamically, trade can depend on trade agreements and trade agreements can depend on trade without creating the logical inconsistencies such dependency would cause in cross-sectional studies.

6.3.1 Building a simple qualitative VAR

Assuming for simplicity's sake that we have only two endogenous variables: trade (X) and regional integration agreements (RIA). Ignoring the endogeneity of trade, a *static* probit model explaining RIAs can be written down using a latent variable $RIA^{\star}$:

$$\begin{aligned} RIA^{\star}_{ij,t} &= \phi_1 X_{ij,t} + x_{ij,t}\, b_1 + c_{1_{ij}} + \varepsilon_{1_{ij,t}} \\ RIA_{ij,t} &= \begin{cases} 0 & \text{if } RIA^{\star}_{ij,t} \leq 0 \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{6.1}$$

$X_{ij,t}$ denotes the (log of the) total trade between countries i and j at time t. $RIA_{ij,t}$ is a dummy variable indicating whether the two countries are members of the same trade agreement at time t and $RIA^{\star}$ is its latent continuous equivalent. $c_{1_{ij}}$ holds a vector of constants/-fixed effects, while $x_{ij,t}$ contains the remaining exogenous explanatory variables. In a probit model, the error term ε_1 is assumed to come from a normal distribution in which variance is normalized to one on order to identify the model.

Baier and Bergstrand (2004a) interpret $RIA^{\star}$ as the *minimal* willingness of both countries to sign an integration agreement. Since both countries have to agree, it is the country with the smallest willingness that will ultimately decide whether or not an agreement is signed. However, this interpretation runs into some problems especially when used in the dynamic setting. Without additional assumptions, there is no guarantee that the minima of two linear functions is itself linear. Moreover, a change in which of the two countries has the smallest willingness would also alter the parameter values. An interpretation that would avoid both problems is if $RIA^{\star}$ is the *average* willingness to sign. The underlying assumption is that countries can compensate each other either monetarily or for example through concessions in other parts of the agreement. A county with a lot to gain from the agreement could in this way try to compensate an unwilling partner, making their average willingness the deciding factor. This would avoid the problems associated with minima, at the cost of introducing bartering to the RIA negotiations.

It should be pointed out that $RIA^{\star}$ is simply a mechanical feature that allows us to write the probit model in a linear way. The meaning we ascribe to it does not alter the parameters of the probit regression, although it does have repercussions for the way in which the theoretical model is translated to the empirical specification. However, as this discussion would lead us too far from the main point of this chapter we will simply refer to $RIA^{\star}$ as the willingness to sign, leaving out whether this is a minimum or an average.

Secondly, a *static* log-linear gravity model that ignores the endogeneity of trade agreements is given by equation 6.2. The error term $\varepsilon_{2_{ij,t}}$ also comes from a normal distribution and has variance σ_2. Using similar control variables x and fixed effects matrix $c_{2,ij}$ we get:

$$X_{ij,t} = \phi_2\, RIA_{ij,t} + x_{ij,t}\, b_2 + c_{2_{ij}} + \varepsilon_{2_{ij,t}} \tag{6.2}$$

To construct a qualitative VAR, the *RIA* dummy in the gravity equation is first replaced by the latent $RIA^{\star}$ from the probit model. Equations 6.2 and 6.1 are then stacked and the endogenous variables are modeled dynamically. Using p lags on each endogenous variable, the reduced form can be written as:

$$\begin{bmatrix} RIA^{\star}_{ij,t} \\ X_{ij,t} \end{bmatrix} = \sum_{k=1}^{p} \Phi^{(k)} \begin{bmatrix} RIA^{\star}_{ij,t-k} \\ X_{ij,t-k} \end{bmatrix} + b\, x'_{ij,t} + c_{ij} + \varepsilon_{ij,t} \tag{6.3}$$

$$RIA_{ij,t} = \begin{cases} 0 & \text{if } RIA^{\star}_{ij,t} \leq 0 \\ 1 & \text{otherwise.} \end{cases} \tag{6.4}$$

$\Phi^{(k)}$ is an $(m \times m)$ matrix holding the parameters on the k^{th} lag of the m endogenous variables. In this simple example m is equal to two, but $X_{ij,t}$ could also be interpreted as a vector containing multiple continuous endogenous variables. The remaining parameters and the error term can be obtained by stacking their counterparts: $b = [b'_1, b'_2]'$, $c_{ij} = \left[c_{1_{ij}}, c_{2_{ij}}\right]'$ and $\varepsilon_{ij,t} = \left[\varepsilon_{1_{ij,t}}, \varepsilon_{2_{ij,t}}\right]'$. The error term $\varepsilon_{ij,t}$ is assumed to come from an independent and identically normal distribution with zero mean and variance matrix Σ. Similar to the probit regression with latent variables, identification of the model requires the assumption that the first diagonal element of Σ is one:

$$\Sigma = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_2 \end{bmatrix}$$

6.3.2 Estimation using Bayesian Gibbs sampling

Using a Bayesian Gibbs sampler allows us to split up the estimation of this system into multiple parts. Instead of having to compute the entire (posterior) probability of all parameters at once, it is separated into various conditional probabilities that are much easier to solve. The Gibbs sampler iteratively draws from those probabilities while conditioning on the values from the previous draws: $a_1 \sim p(a|b_1)$, $b_2 \sim p(b|a_1)$, $a_2 \sim p(a|b_2)$, etc. After a certain number of draws, these draws will have converged to the unconditional posterior and the remaining draws can be used to reconstruct the distribution of the parameters (Koop and Korobilis, 2009).

To simplify the notation used in the remainder of this section, the parameters of the qualitative VAR are condensed to the variance Σ and the parameter coefficients $\Theta = \{c_{ij}, b, \Phi\}$ where $\Phi = \{\Phi^{(1)}, \cdots, \Phi^{(p)}\}$.

If the latent variable $RIA^{\star}_{ij,t}$ were known, equation 6.3 could be estimated using seemingly unrelated regression techniques. However, computing and drawing values for the latent variable conditional on the parameters in equation 6.3 (Θ and Σ) is less straightforward. The mean and standard deviation of $RIA^{\star}_{ij,t}$ depend on past and future values of the endogenous variables, as well as the current values of the exogenous variables. However, as Dueker (2005b) noted, a simple rewrite of this model reveals a state-space model which can be estimated and drawn from using a modified Kalman filter (cf. infra, section 6.3.2). In addition to the computational convenience this offers, the multi-move sampling technique also ensures a faster convergence.

By imposing an independent normal-Wishart prior on Θ and Σ, the conditional posterior distributions remain relatively simple. Throughout this chapter, we used an uninformative prior on Σ, combined with a Minisota prior on Θ. The Minisota prior allows for prior shrinkage, exponentially decreasing the weight of the parameters on higher lags. This helps ensure that the Gibbs sampler converges even when the number of endogenous variables and lags increases (Koop and Korobilis, 2009).

The matrix c_{ij} can be adjusted to estimate a wide range of models, including sender and target fixed effects that control for (time-invariant) multilateral resistance terms. In a probit model, the incidental parameter problem cannot be circumvented by using demeaned variables. As a result, the fixed effects can only be estimated by including a large number of dummy variables (Egger et al., 2011). Following Guimarães and Portugal (2009), the estimation of the dummies is separated from the other variables in Θ, keeping the size of the matrix that needs to be inverted under control.

Figure 6.1 summarizes the different loops in the Gibbs sampler. From left to right, it provides an overview of how the Gibbs sampler separates the posterior distribution of θ and Σ into conditional probabilities. Step A shows how Dueker (2005a,b) first split up the posterior by introducing the latent variable $RIA^{\star}$. The next section describes how $RIA^{\star}$ can be computed and drawn from if we know what Θ and Σ are. Step B and C illustrate how those parameters can be drawn conditional on $RIA^{\star}$. Appendix 6.A lists the probability distributions of each step, but for an exhaustive overview we refer the reader to Koop and Korobilis (2009) and Guimarães and Portugal (2009).

Figure 6.1: Structure of the Gibbs sampler algorithm

$$\theta,\Sigma|RIA,X,x \underset{A}{\Rightarrow} \begin{cases} RIA^{\star}|RIA,\Theta,\Sigma,X,x \sim N \\ \Theta,\Sigma|RIA^{\star},X,x \quad \underset{B}{\Rightarrow} \begin{cases} \Sigma|\Theta,RIA^{\star},X,x \sim iW \\ \Theta|\Sigma,RIA^{\star},X,x \underset{C}{\Rightarrow} \begin{cases} b,\Phi|\Sigma,RIA^{\star},X,x,c_{ij} \sim N \\ c_{ij}|\Sigma,RIA^{\star},X,x,\Phi,b \sim N \end{cases} \end{cases} \end{cases}$$

$$\begin{array}{rcl} N & : & \text{Normal distribution} \\ iW & : & \text{inverse Wishart distribution} \\ \underset{A}{\Rightarrow} & : & \text{Dueker (2005a,b)} \\ \underset{B}{\Rightarrow} & : & \text{Koop and Korobilis (2009)} \\ \underset{C}{\Rightarrow} & : & \text{Guimarães and Portugal (2009)} \end{array}$$

The conditional distribution of the latent variable $RIA^{\star}$

The final step of the Gibbs sampler computes and draws from the distribution of $RIA^{\star}$, conditional on the parameters of the qualitative VAR. In a static probit model this can be solved by drawing from a truncated normal distribution to ensure the values of $RIA^{\star}$ are positive when an agreement is signed and vice versa. In the qualitative VAR on the other hand, the dynamics make it so that the distribution of $RIA^{\star}$ at moment t will depend on the previous values and will in turn influence future values. However, instead of having to compute this dependence over p lags and estimate $RIA^{\star}$ for the entire time-period, the qualitative VAR can be rewritten into a state-space model which can solved observation by observation.

A state-space model is built around two equations that define the behavior of an unknown, to-be-estimated state vector. The state equation (equation 6.5) describes the change in the state vector S_t over time: the way in which it depends on its previous values (μ and F) and how big the changes in each period can be (ν_1). Secondly, the measurement equation (equation 6.6) specifies how this state-vector in turn is related to a number of observed variables (X_t). Specifically, it states how the observed variables are scaled (H) and what their reliability is (ν_2). The error terms ν_1 and ν_2 are assumed to be normally distributed.

$$
\begin{aligned}
S_t &= \mu + F S_{t-1} + \nu_{1,t} && (6.5)\\
X_t &= H\, S_t + \nu_{2,t} && (6.6)
\end{aligned}
$$

The Kalman filter and smoother algorithms can be used to compute the distribution of the state vector at each point in time. The strength of these algorithms lies in the fact that they do this iteratively which significantly reduces the computational burden. In each step they use the state equation to predict the current value of S_t based on the past (Kalman filter) or future (Kalman smoother) estimates of S. This prediction is then updated using the information in X_t whose scaling and reliability is determined by the measurement equation (Kim and Nelson, 1999).

Applying this logic to the qualitative VAR model, the willingness to sign ($RIA^\star$) is the unknown state while the information in $RIA_{ij,t}$ and $X_{ij,t}$ serves as the observed measurements. To rewrite equation 6.3 as a state-space model, the vector of endogenous variables is first summarized as a $(m \times 1)$ vector $Y_{ij,t} = \left[RIA^\star_{ij,t}, X_{ij,t}\right]'$. The state variable is subsequently obtained by stacking p lags of this vector, $S_t = [Y'_{ij,t}, \cdots, Y'_{ij,t-p+1}]'$, resulting in the following model:

$$
\begin{bmatrix} Y_{ij,t} \\ Y_{ij,t-1} \\ \vdots \\ Y_{ij,t-p+1} \end{bmatrix}
=
\begin{bmatrix} c_{ij,t} + b\, x'_{ij,t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}
+
\begin{bmatrix} \Phi^{(1)} & \Phi^{(2)} & \dots & \Phi^{(p)} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix}
\begin{bmatrix} Y_{ij,t-1} \\ Y_{ij,t-2} \\ \vdots \\ Y_{ij,t-p} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{ij,t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}
\qquad (6.7)
$$

$$
X_{ij,t} = \begin{bmatrix} 0_{1\times m-1} & I_{m-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}
\begin{bmatrix} Y_{ij,t} \\ Y_{ij,t-1} \\ \vdots \\ Y_{ij,t-p+1} \end{bmatrix}
\qquad (6.8)
$$

The first row of the state equation (6.7) simply repeats the qualitative VAR model (equation 6.3). $c_{ij} + b\,x'_{ij,t}$ is a simple -albeit time-varying- scalar, since this step of the Gibbs sampler algorithm is conditional on the parameter values Θ and Σ. The measurement equation (6.8) establishes the relation between $Y_{ij,t}$ and the continuous endogenous variable(s) $X_{ij,t}$. Without an error term in the measurement equation, only the first element of the state variable can vary in each draw: $RIA^{\star}_{ij,t}$. The values of the other endogenous variables are kept fixed.

The main difference with a standard state-space model is that the error term is not multivariate normally distributed. Similar to the probit model, the error term has to be drawn from a truncated normal distribution to ensure that $RIA^{\star}$ is positive when a RIA is signed. This means that the expected value and standard deviation of $\varepsilon_{ij,t}$ changes depending on whether or not a trade agreement has been signed. Appendix 6.A gives an overview of how this affects the Kalman filter and smoother algorithms.

6.3.3 Identifying the structural model

Because this chapter is intended more as a proof of concept of using a qualitative VAR in the analysis of trade flows, the identification of the structural model has purposefully been kept simple. A Cholesky decomposition is used to impose a strict ordering in the timing of each variable. Other possible identification methods are discussed in the extensions (section 6.6). It should be mentioned that the choice of identification strategy will only affect the structural impulse response functions. The average treatment effects on the other hand are computed using the reduced model's parameters.

While trade agreements are assumed to be able to immediately affect trade, the willingness to close trade agreements adjusts more slowly.[6] This reflects the fact that the negotiation of trade agreements takes time. When added as an endogenous variable, the remaining variables are ordered as: 1) RIA; 2) trade; 3) capital-labor ratio;

[6] This falls along the lines of the restriction used in Baier and Bergstrand (2002, section VII-A) that ensures the logical consistency of the cross-sectional model.

4) difference in GDP; 5) average GDP. The cholesky decomposition imposes that each variable has no immediate effect on those preceding it, but can be contemporaneously affected by them.

6.4 Data

The baseline model uses a simple dummy indicator that captures whether or not two countries are currently members of the same trade agreement (*RIA*). This variable was composed using the information in the WTO's Regional Integration Agreements Information System and the United Nations University's Comparative Regional Integration Studies electronic platform: the Regional Integration Knowledge System. Both databases combined provided information on 251 agreements covering 205 countries from 1950 to 2015. Agreements between a customs union and another country were ascribed to all members of the customs union at that time. The complete list can be found in appendix 6.B.

Following Baier and Bergstrand (2004a), the trade agreements variable was defined per country-pair. This gave a total of $\frac{205\times 204}{2} \approx 20,000$ country couples and 850,000 observations. However, when combined with the availability in trade data and discarding zero-trade flows about 275,000 observations are left. Trade flows were measured as the sum of the logs of exports and imports.[7] For now, zero trade flows were ignored, but a solution to this problem in the line of Egger et al. (2011) is discussed in the extensions (section 6.6). The other endogenous variables are the GDPs of both countries and the difference in their capital-labor ratio (*DKL*). Bilateral trade data was supplied by the IMF's Direction of Trade Statistics while the Penn world tables 8.0 provided information on GDP, population and capital (Feenstra et al., 2013).

Information on distance, population and capital was used to create variables expressing the remoteness of two countries relative to the other countries on their

[7] This avoids the silver medal mistake of gravity equations which is to take the log of the sum (Baldwin and Taglioni, 2006).

continent (*remote*) and the extent to which their capital-labor ratio differs from that of the rest of the world (*DROWKL*). Both variables were computed as described in Baier and Bergstrand (2004a). The availability of the capital-labor data was similar to that of GDP allowing us to use both (unlike for example Egger et al., 2011).

Proxies for ice-berg type trade costs were also included as control variables, most of which came from CEPII's gravity dataset (Head et al., 2010; Head and Mayer, 2013). These include the log of (population-weighted) *distance* and a series of dummies indicating whether two countries neighbor another (*contiguity*), whether one country was once a colony of the other (*colony*), whether they were once colonized by the same country (*common colony*), whether they share an ethnographic *language* and whether one of the countries is *landlocked*. Finally, following Egger et al. (2011) a number of political variables from the polity IV project were included (Marshall et al., 2014). *Autocracy*, political competition (*pol. comp.*) and *durability* measure the absolute distance of the country-couple in terms of those political characteristics. Appendix 6.C provides summary statistics.

6.5 Results

Similar to the identification of the structural model, the model specification is kept simple. The starting point for the gravity equation is a log-linear version of the one used in Anderson and van Wincoop (2003), while the probit model's specification is based on Baier and Bergstrand (2004a). The gravity equation includes country-fixed effects to control for multilateral resistance terms, but unlike Baldwin and Taglioni (2006); Baier and Bergstrand (2007) or Head and Mayer (2013) they are kept constant over time.[8] A further simplification is that the same exogenous control variables are used in both equation. The issue of making the model specification more consistent with the trade theory is revisited in section 6.6.

Both equations are adjusted to the VAR framework by including the endogenous

[8]Egger and Nigai (2015) note that the country-year fixed effects are correlated with the error term and as a result still produce biased results.

variables ($Y_{ij,t}$) dynamically. In other words, as opposed to static models that try to estimate the long-term equilibrium relation, the focus is shifted to the adjustment to the long-term equilibrium. This gives rise to two models: a *limited* model where only trade and RIAs are endogenous and the *full* model where the GDP and capital labor ratios are also modeled endogenously. In both cases, the reduced form of the qualitative VAR can be written as:

$$Y^{\star}_{ij,t} = \sum_{k=1}^{p} \Phi^{(k)}\, Y_{ij,t-k} + b\, x'_{ij,t} + c_i + c_j + \varepsilon_{ij,t} \tag{6.9}$$

with $x_{ij,t}$ a vector of control variables, c_i and c_j country fixed effects and $\varepsilon_{ij,t}$ the normally distributed error term with variance-covariance matrix Σ.

6.5.1 Limited model

In the limited model $Y_{ij,t}$ is equal to $[RIA^{\star}_{ij,t}, X_{ij,t}]'$. The exogenous variables $x_{ij,t}$ control for the country size and relative factor endowments by including the log of the GDPs of both countries and the difference in their capital labor ratios (*DKL*) in addition to the other control variables listed in the previous section. The GDPs were labeled such that country i is on average larger than country j throughout the period of the study.[9]

The parameter values of the reduced model are listed in table 6.1, however these cannot be used directly to study the effects of a change as this would ignore the dynamics of the system. To take these into account, the Cholesky decomposition is first used to transform the model into its structural equivalent as detailed in section 6.3.3. The structural parameters are subsequently used to compute the impulse response functions (irf) shown in figure 6.2. The irf show the change in the endogenous variables in response to a temporary shock (or impulse) in each of the endogenous variables. These shocks, indicated between brackets, happen at moment $t = 0$ and correspond to one standard deviation in the shocked variable. The x-axis shows the number of years since the shock and the y-axis shows the resulting

[9] This only altered the labels on the variables without affecting the selection of country-pairs.

change in the value of the variable in question. Finally, the 90% confidence intervals are indicated by the blue dotted lines. We will use coordinates to refer to an individual response function by counting the number of rows and columns starting from the top left corner (cf. matrices).

Table 6.1: Reduced parameter values of the limited model - World

	RIA⋆	st.e.	X	st.e.
L1.RIA⋆	0.0456[a]	(0.0051)	0.0051[a]	(0.0010)
L2.RIA⋆	−0.0036[a]	(0.0013)	0.0008	(0.0007)
L1.X	0.0788[a]	(0.0077)	0.6015[a]	(0.0023)
L2.X	0.0548[a]	(0.0067)	0.2430[a]	(0.0022)
GDP_i	0.5394[a]	(0.0206)	0.1825[a]	(0.0039)
GDP_j	0.0478[a]	(0.0072)	0.1641[a]	(0.0035)
DKL	−0.1210[a]	(0.0073)	−0.0082[a]	(0.0025)
DROWKL	−0.7203[a]	(0.0408)	0.0255[c]	(0.0159)
Distance	−1.0381[a]	(0.0220)	−0.1932[a]	(0.0044)
Contiguous	−0.4161[a]	(0.0310)	0.1122[a]	(0.0114)
Landlocked	0.0508	(0.0563)	−0.0899[a]	(0.0164)
Remote	0.0050[b]	(0.0022)	0.0015[b]	(0.0007)
Colony	0.061	(0.0489)	0.2466[a]	(0.0162)
Common colony	−0.2591[a]	(0.0292)	0.1289[a]	(0.0109)
Language	0.0708[a]	(0.0161)	0.0748[a]	(0.0069)
WTO	0.1973[a]	(0.0271)	−0.0018	(0.0074)
Autocracy	0.0416[a]	(0.0034)	0.0012	(0.0012)
Pol. comp	−0.0404[a]	(0.0033)	−0.0014	(0.0011)
Durability	−0.0038[a]	(0.0003)	−0.0006[a]	(0.0001)
nObs	167410		167410	
Fixed effects	sender & target		sender & target	

Reduced parameter estimates of the limited, worldwide qualitative VAR model with two lags. Standard errors between brackets. [a], [b] and [c] indicate significance at the 1%, 5% and 10% level.

For example, the top right panel has coordinates [1,2]. The title $RIA^\star$ (X) indicates that it plots the change in the value of the latent variable $RIA^\star$ in response to a shock in the log of trade of one standard deviation. It reveals that an increase in trade will significantly raise the willingness to enter into a RIA, corroborating the natural trading partner hypothesis of Krugman (1995). Secondly, panel [2,1] shows that a shock to $RIA^\star$ will significantly raise bilateral trade. However, while the irf can reveal the sign and significance of the effect of trade agreements on trade, they cannot be used to measure the size of the effect. The reason is that it is not clear whether a shock of one standard deviation in the *willingness* to sign an agreement

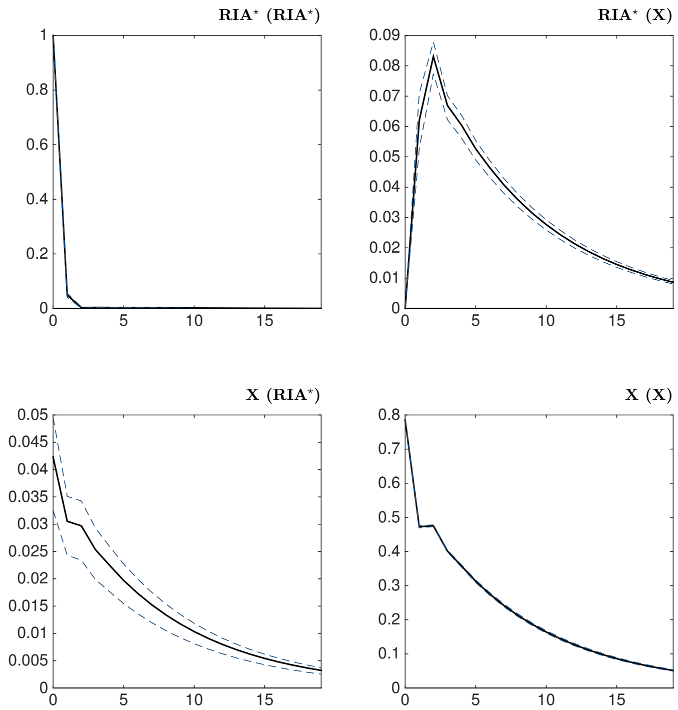


Figure 6.2: Structural impulse response functions of the limited model - World
Responses in trade and the willingness to form a trade agreement following a shock of one standard deviation in the impulse variable (between parentheses). The x-axis shows the number of years since the shock (at $t = 0$). 90% confidence intervals are indicated by the blue interrupted lines.

would actually result in an agreement being signed (cf. section 6.5.3).[10]

With a few exceptions the behavior of most control variables falls within expectations (table 6.1). The long-term parameter on GDP and distance in the gravity equation are (slightly) higher than one, but lie within the bounds of what is found in other studies (Head and Mayer, 2013): $\bar{\beta}_{GDP_i} = \frac{0.183}{1-(0.602+0.243)} \approx 1.17$; $\bar{\beta}_{GDP_j} \approx 1.01$ and $\bar{\beta}_{\text{Distance}} \approx -1.24$. The negative coefficient on *DKL* in the $RIA^\star$ equation does not match with the findings of Baier and Bergstrand (2004a), but for example Magee (2003) and Márquez-Ramos et al. (2011) found similar signs in their probit regression.[11] The negative coefficients on DROWKL and contiguity are unexpected, but are counteracted through their effect on trade. Moreover, they disappear when DKL

[10]The sign and significance can nevertheless be identified because the values of $RIA^\star$ are determined by the actual value of the *RIA* dummy (equation 6.4).

[11]Magee (2003) connects the negative coefficient on *DKL* to the political economy argument of Levy (1997) that agreements are more likely to form between homogenous countries. A small difference in capital-labor ratios indicates a similar economic structure, which raises the likelihood that an agreement can be reached.

and GDP are considered endogenous (cf. infra). In contrast with the instruments used in Egger et al. (2011), colonial history is an inconsistent predictor of trade agreements once the level of trade is controlled for: *colony* is insignificant and while common colonial history is significant it changes sign in the full model. This lends further weight against the practice of estimating the effect of trade agreements through an instrumental variable approach (Baier and Bergstrand, 2004b). The remaining political variables also perform inconsistently. Only the similarity in terms of political competition will consistently positively affect the willingness to sign.

6.5.2 Full model

In contrast with the limited model, the full model uses the average and difference in the log of the GDPs. This is done so that all endogenous variables vary on three dimensions (sender-target-year), as opposed to two dimensions when the level of GDP of both countries is entered separately (country-year). Combining data in different dimensions would otherwise create problems when stacking data on different countries/country-pairs.[12] Using the averages and differences is how GDP is typically modeled in the probit regression (Baier and Bergstrand, 2004a). The implication for the gravity model is that the same coefficient is imposed on both GDPs, an assumption that is consistent with the theory and can be found throughout the literature (e.g. Baldwin and Taglioni, 2006; Baier and Bergstrand, 2007).

When the difference in capital labor ratios is also considered endogenous, $Y_{ij,t}$ is equal to $[RIA^{\star}_{ij,t}, X_{ij,t}, DKL_{ij,t}, \text{GDP}diff_{ij,t}, \text{GDP}av_{ij,t}]'$ in the full model. The control variables in $x_{ij,t}$ remain the same (except for those that are now treated as endogenous) and country fixed effects are included in all equations.

The model was first run for the entire world (figure 6.3) after which the estimation repeated for only European countries (figure 6.4) and African countries (figure 6.5). The three figures paint a very similar picture overall, but the sign and significance

[12]The literature on global VARs deals with variables of different dimensions (e.g. Pesaran et al., 2004), but in this framework this would imply estimating a model with more than 40,000 equations as each country-couple's trade and willingness to sign RIAs would have to be estimated simultaneously.

of some irf can change depending on the region studied. Overall the interaction between trade and trade agreements is not altered when GDP and DKL are considered endogenous. The effect of a shock to the willingness to sign on trade remains positive when DKL and GDP are considered endogenous (panel [2,1]). The effect of a shock to trade on $RIA^{\star}$ is also positive and while it is barely significant for the world, it is strongly significant in both subsamples (panel [1,2]). Furthermore, an increase in the willingness to sign will decreases the difference in GDP and increase the average GDP in all samples (panels [4,1] and [5,1]). A shock to trade on the other hand will increase average GDP (panel [5,2]), but its effect on DKL and the GDP_{diff} changes depending on the estimation sample (panels [3,2] and [4,2]). As was the case in the limited model, an increase in DKL will lower the willingness to sign (panel [1,3]). Worldwide it will also lower trade but the opposite is true in the European and African subsamples (panel [2,3]). The effect of an increase in the difference in GDP on $RIA^{\star}$ is ambiguous, but it will decrease trade (panels [1,4] and [2,4]). Finally an increase in the average GDP will increase both trade and the likelihood of signing an agreement, but this is not always significant (panels [1,5] and [2,5]).

6.5.3 Assessing the effect of trade agreements on trade

The impulse response functions shown earlier help give an insight into the sign and long run dynamics of the effect of trade on the willingness to join a regional integration agreement. However, there is a difference between the effect of "a rise in the willingness to sign" on trade and the effect of "signing" a trade agreement. Similar to the interpretation of the estimation results of a probit model, the parameter values are not equal to the marginal effect on the dependent variable. While the impulse response functions can show the importance of taking the dynamics and endogeneity into account, they cannot be used to estimate the average treatment effect of signing a trade agreement.

The average treatment effect of RIA on trade can be expressed as $ATE = E(X|.,RIA =$

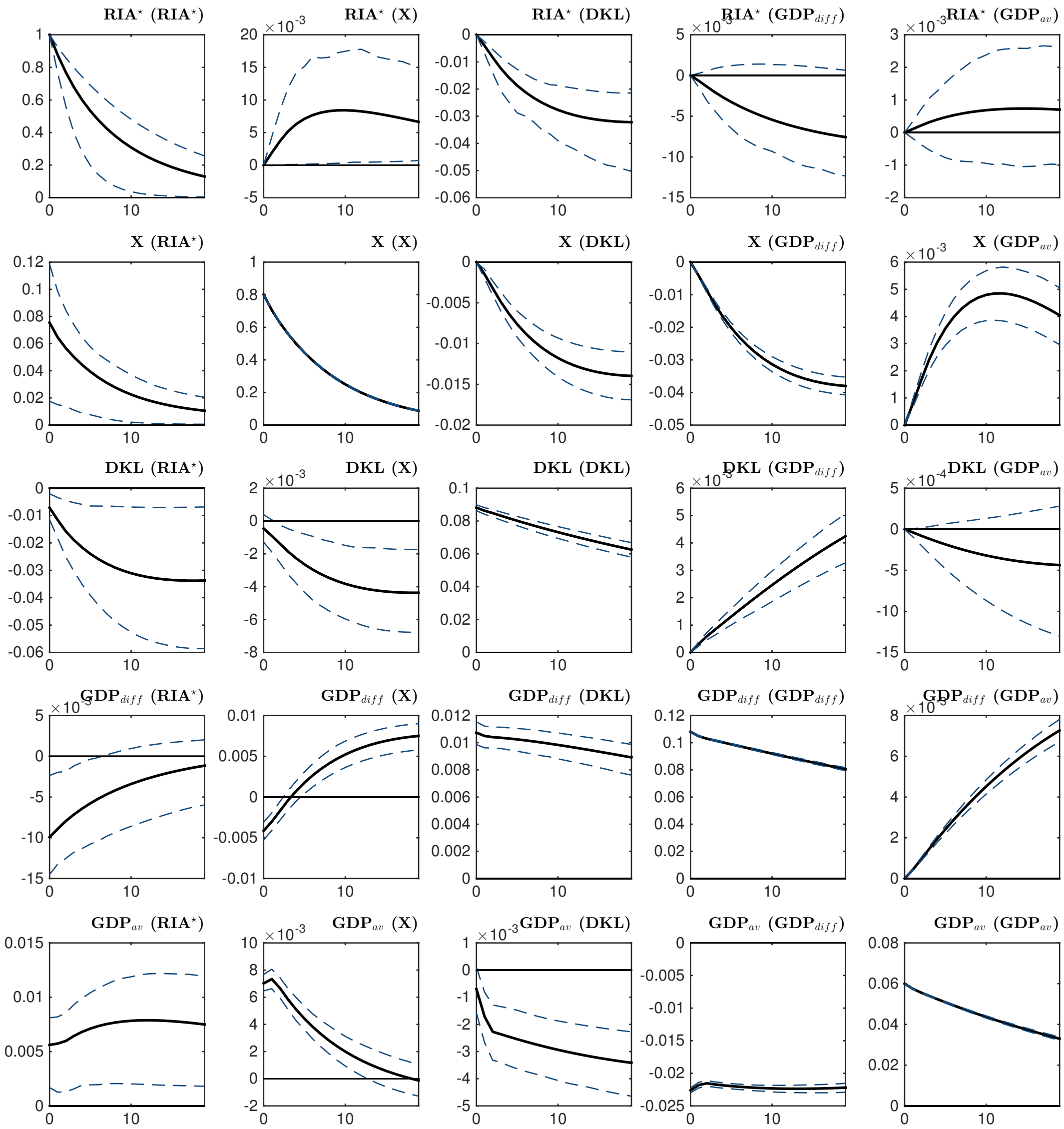


Figure 6.3: Structural impulse response functions of the full model - World
Responses in trade, the willingness to form a RIA, average and difference in GDP and the capital-labor ratio to a shock of one standard deviation in the impulse variable (between parentheses). The x-axis shows the number of years since the shock (at $t = 0$). 90% confidence intervals are indicated by the blue interrupted lines. The variables are listed in the order of the Cholesky decomposition.

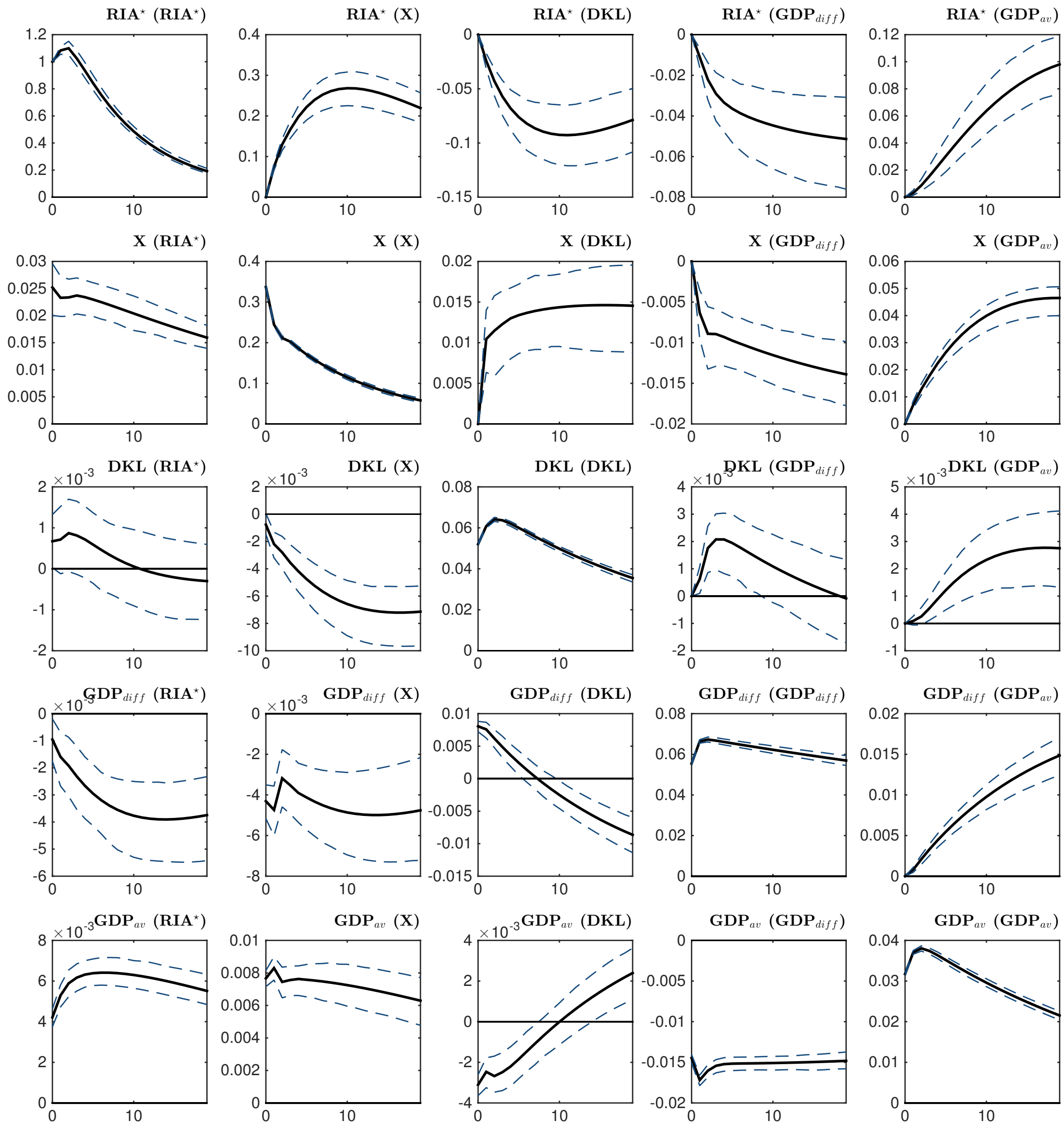


Figure 6.4: Structural impulse response functions of the full model - Europe
Responses in the willingness to form a RIA, trade, average and difference in GDP and the capital-labor ratio to a shock of one standard deviation in the impulse variable (between parentheses). The x-axis shows the number of years since the shock (at $t = 0$). 90% confidence intervals are indicated by the blue interrupted lines. The variables are listed in the order of the Cholesky decomposition.

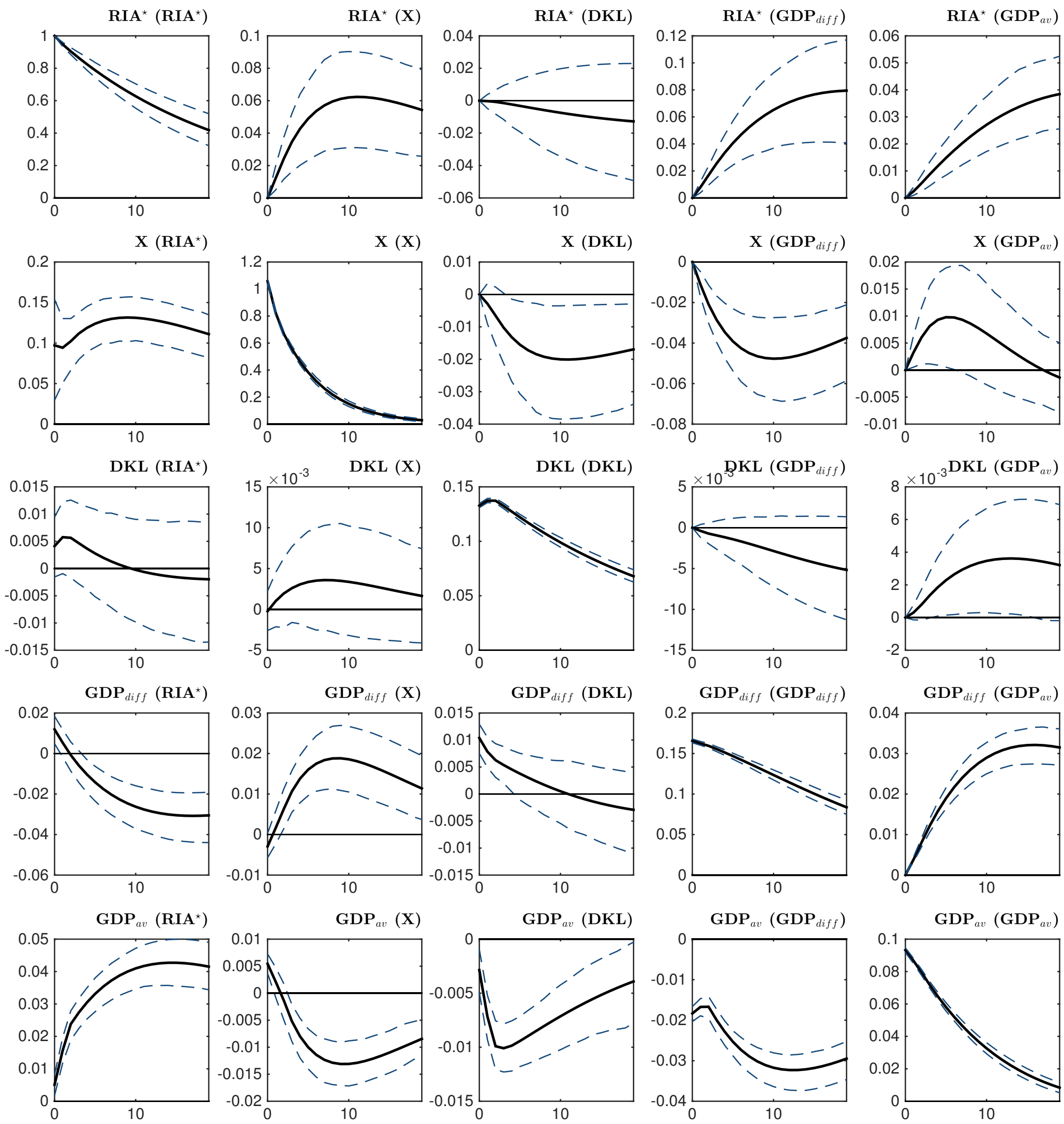


Figure 6.5: Structural impulse response functions of the full model - Africa
Responses in trade, the willingness to form a RIA, average and difference in GDP and the capital-labor ratio to a shock in the impulse variable of one standard deviation (between parentheses). The x-axis shows the number of years since the shock (at $t = 0$). 90% confidence intervals are indicated by the blue interrupted lines. The variables are listed in the order of the Cholesky decomposition.

$1) - E(X|.,RIA = 0)$. The difficulty assessing the treatment effect is identifying the right counterfactual. Either a country-pair signed an agreement and what trade would be without an agreement is unknown, or vice versa. The dummies that traditionally have been used in gravity equations have been shown to lead to severe parameter instability, even when controlling for endogeneity. Their sign, size and significance changes depending on the study, methodology and even the included control variables (e.g. Magee, 2003). Instead, Baier and Bergstrand (2009) used a non-parametric matching technique to find existing county-couples with similar characteristics but without a trade agreement. Egger et al. (2011) used the estimated parameters on trade and trade agreements to generate the appropriate counterfactual for each country-couple.

The approach we suggest is similar to that of Egger et al. (2011). Using the business-cycle filter from Dueker and Nelson (2006) it is possible to generate values of trade conditional on any value of $RIA^{\star}$. By ensuring that the willingness to sign is never greater than zero, we can impose that no trade agreement was signed. The counterfactuals are generated by reversing the roles of the variables in the state-space model described in section 6.3.2. $RIA^{\star}$ is fixed at $\bar{r}$ while new values of the other endogenous variables are computed and drawn ($\tilde{X}$). Substituting $\tilde{Y}_{ij,t} = \left[RIA^{\star}_{ij,t}, \tilde{X}_{ij,t}\right]'$ results in a similar state equation, but changes the measurement equation.

$$\begin{bmatrix} \tilde{Y}_{ij,t} \\ \tilde{Y}_{ij,t-1} \\ \vdots \\ \tilde{Y}_{ij,t-p+1} \end{bmatrix} = \begin{bmatrix} c_{ij,t} + b\,x'_{ij,t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \Phi^{(1)} & \Phi^{(2)} & \cdots & \Phi^{(p)} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{Y}_{ij,t-1} \\ \tilde{Y}_{ij,t-2} \\ \vdots \\ \tilde{Y}_{ij,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij,t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{6.10}$$

$$\bar{r} = \begin{bmatrix} 1 & 0_{1\times m-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{Y}_{ij,t} \\ \tilde{Y}_{ij,t-1} \\ \vdots \\ \tilde{Y}_{ij,t-p+1} \end{bmatrix} \tag{6.11}$$

The model specified in equations 6.10 and 6.11 treats the latent integration agreement variable as the only observed data. However, this approach can be further augmented to also take historical data into account. In that case, the counterfactual will try to follow historical data to the extent that it corresponds with an unchanged willingness to sign a RIA. Incorporating the historical aspect becomes especially interesting as more variables (for example GDPs and capital-labor ratios) are modeled as endogenous. To generate values of trade that fall in between these two cases, Dueker (2005a) proposes the following measurement equation:

$$\begin{bmatrix} \bar{r} \\ \alpha X_{ij,t} \end{bmatrix} = \begin{bmatrix} 1 & 0_{1\times m-1} & \mathbf{0} & \cdots & \mathbf{0} \\ 0_{m-1\times 1} & \alpha I_{m-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{Y}_{ij,t} \\ \tilde{Y}_{ij,t-1} \\ \vdots \\ \tilde{Y}_{ij,t-p+1} \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha \eta_{ij,t} \end{bmatrix} \tag{6.12}$$

If the smoothing parameter α is zero, equation 6.12 is the same as 6.11. However, as α grows the counterfactual will increasingly reflect the historical data. The error term $\eta_{ij,t} = X_{ij,t} - \tilde{X}_{ij,t}$ is normally distributed with mean zero and variance matrix Ω. The latter can be drawn from an inverse Wishart distribution in the same way as Σ (Dueker and Nelson, 2006). To further reduce the informational value of the historical series, their values were set to missing whenever a trade agreement was signed.

To illustrate, figure 6.6 plots both actual values and the computed counterfactual for the bilateral trade between Mexico and the United States. The black lines show the actual values of the endogenous variables, while the counterfactual and its 90% confidence interval are indicated by the blue interrupted and dotted lines. In addition to the dummy indicating whether or not a trade agreement was signed (*RIA*), the top panel also shows the willingness to sign trade agreements ($RIA^{\star}$) and its 90% confidence interval. This shows quite clearly that as an agreement is signed, $RIA^{\star}$ changes from negative to positive. Also plotted in the top panel is the value of $RIA^{\star}$ that was used to compute the counterfactual. $\bar{r}$ follows $RIA^{\star}$ until a trade agreement is signed after which is set to zero. This means that the counterfactuals are generated under the highest possible willingness to sign that still corresponds to no agreement being signed. In this way, they correspond to a lower bound on the effect of the RIA.

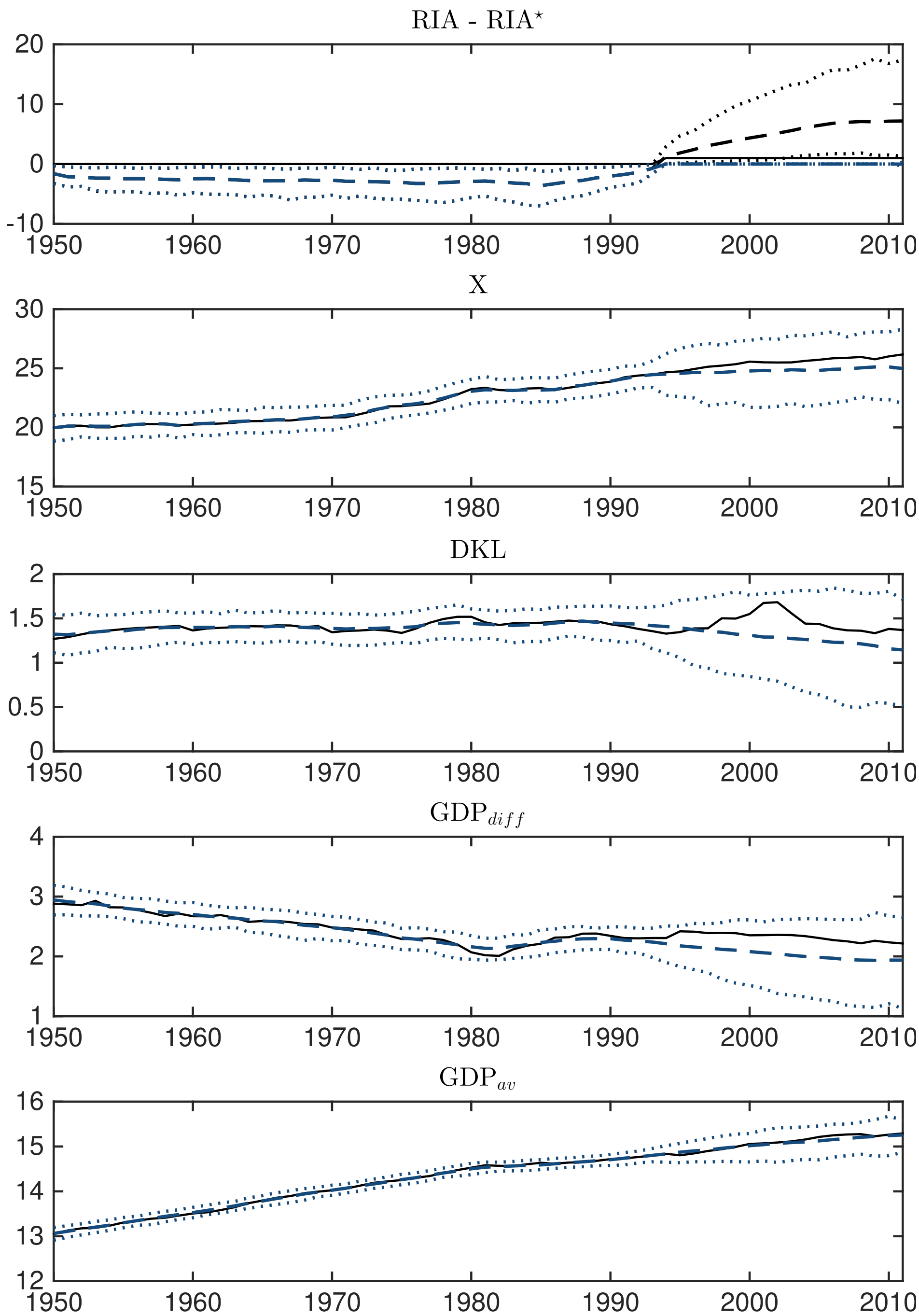


Figure 6.6: Counterfactual flows for Mexico-United States (full model)
Estimated treatment effect of trade agreements on the Mexico-United States willingness to form trade agreements, log trade, difference in captial-labor ratio, difference in log GDP and average of log GDP. Actual values are indicted by the full black line. The counterfactual and its 90% confidence interval are indicated by the blue interrupted and blue dotted lines, respectively.

The remaining panels of figure 6.6 show what trade, the difference in the capital labor ratios and the difference in, and average of GDP would have been if Mexico's and the United States' willingness to sign an agreement never rose above zero. Almost immediately after signing the agreement, the counterfactual level of trade (in logs) starts to diverge from the actual level and the estimated increase from signing a trade agreement only becomes bigger over time. However, this difference remains within the 90% confidence bounds. The counterfactual difference in the capital labor ratios and the GDPs is lower than the actual values, indicating that NAFTA led both countries to diverge. At the same time, the effect on the average GDP is negligible.

The average treatment effect is computed from the individual counterfactual flows. The percentage difference between real values and counterfactual was averaged starting from the moment a trade agreement was signed. If y^s_{ij} year in which the country-couple ij signed a trade agreement, the average treatment effect of an agreement after τ years is:

$$ATE_\tau = \text{mean}_{ij}\left(X_{ij,t-y^s_{ij}+\tau} - \tilde{X}_{ij,t-y^s_{ij}+\tau}\right) \tag{6.13}$$

where the mean is taken over all country couples that have signed a trade agreement at least τ years ago, with the exception of those that entered the dataset with an active trade agreement.

Figure 6.7 plots the worldwide average treatment effect of a trade agreement over time for all endogenous variables. It shows that the average percentage increase in trade is 10% in the first year, 40% after 5 years and 50% after 10 years. It subsequently rises slowly to 80% after 35 years. However, this estimate is based on fewer country-couples, causing the width of the confidence interval to increase strongly. The bottom panel shows that the number of country couples decreases steadily from around a thousand in the first five years to less than a hundred for the last five years. Overall, integration agreements have lead to a convergence of the capital labor ratios, but a divergence in terms of GDPs. The effect on the average

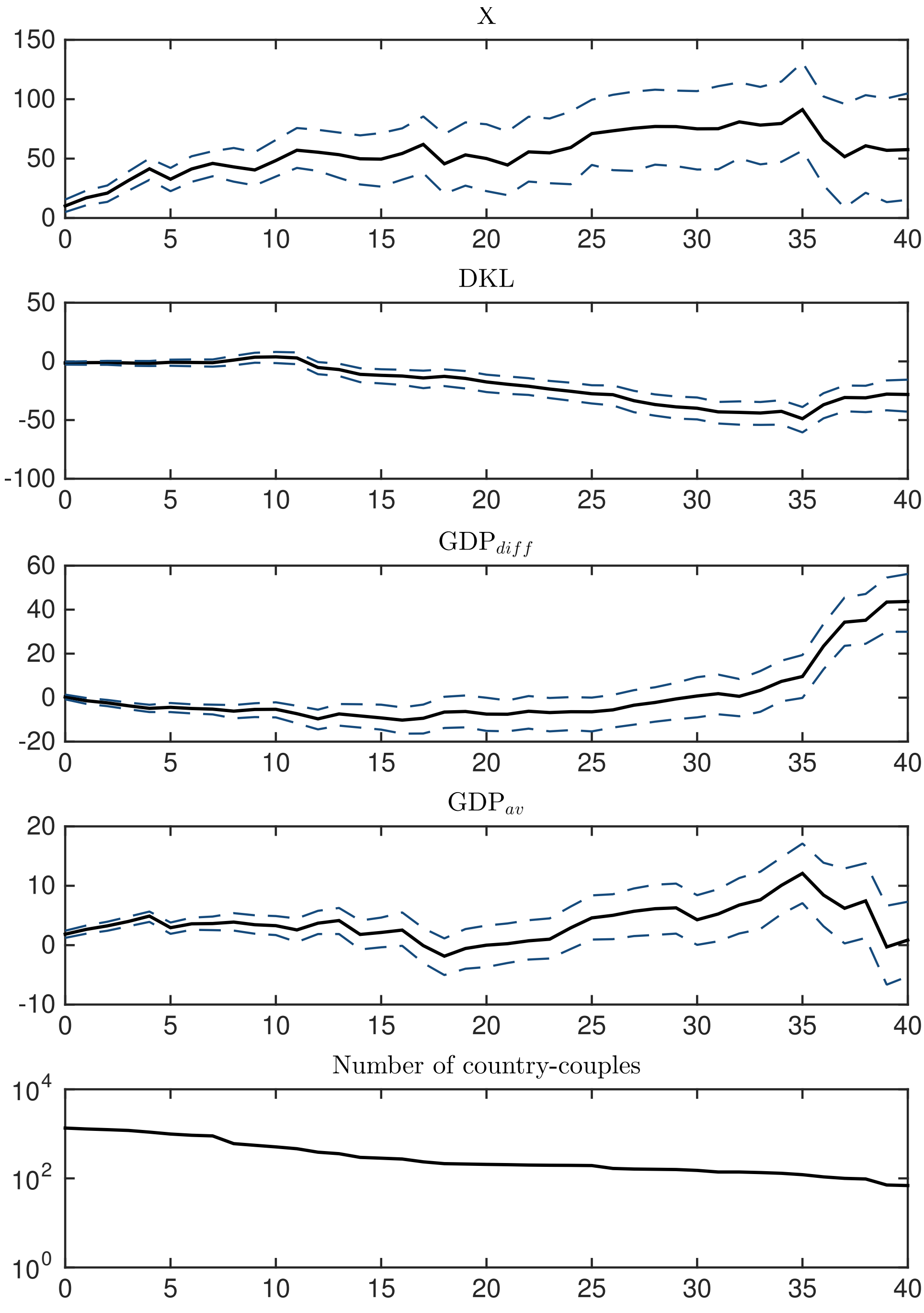


Figure 6.7: Average treatment effects in percentage terms (full model)
Average treatment effect of trade agreements on trade, the difference in capital-labor ratios, difference in GDP and average GDP. 90% onfidence interval indicated by the interrupted blue lines.

GDP is small to zero in the first 20 years. It subsequently starts to increase to about 10% after 35 years. In combination with the increase in the difference in GDP, this seems to indicate that the increase in GDP is one-sided and possibly even at the expense of the growth of the partner country. However, the effects after 20 years might simply be a characteristic of the smaller group of country-couples that have had an agreement for this long.

6.6 Extensions

In contrast with the dynamic nature of the qualitative VAR, the theories underlying the gravity model and the formation of trade agreements are essentially static. The model estimated in equation 6.9 has naively translated the econometric specification to a dynamic setting, while ignoring the underlying theoretical models. A first important extension would be to ensure that the the dynamic equations are theoretically sound, especially if they are used to generate counterfactuals. A first extension to the model would be to explicitly incorporate the time-varying multilateral resistance terms as latent variables, allowing us to control for the indirect, general-equilibrium effect of trade agreements (cf. Egger et al., 2011).
Secondly, in the current model the equations on the endogenous variables are treated the same and include the same control variables. Moreover, the endogenous variables have all been allowed to directly influence one another. However, through a simple change in the priors the direct influence of for example the average GDP on the difference in capital labor ratio could be removed. These and other restrictions, including the number of lags, could subsequently be tested using Bayesian model selection techniques (Koop, 2003).
A third simplification concerns the way integration agreements have been incorporated. The $RIA_{ij,t}$ dummy considered all integration agreements equal, overlooking the vast differences between them. The qualitative VAR model can be extended relatively easily to incorporate agreements of different depth by extending the probit model to an ordered probit model (Dueker, 2005b). The difficulty would lie in the

categorization of the integration agreements.[13] Wu (2006) for example constructed a database dividing agreements into preferential trade agreements, customs unions, common markets, and economic unions. Alternatively, Kohl et al. (2014) deconstructed the depth of 296 trade agreements, checking whether they made any legally enforceable restrictions in 17 trade-related policy domains. This dataset would allow the construction of a index of the depth of an agreement in a way that did not depend so strongly on a one-track, EU-dominated view of regional integration.

Additionally, while the identification of the structural model does not affect the average treatment effects, its influence on the impulse response functions should be checked. A first robustness test would be to impose a different ordering of the variables in the Cholesky decomposition. Other possible ways of identifying the structural model include sign restrictions. The latter would also allow us to look at the timing of the effects of trade agreements in addition to strengthening the link between the qual VAR with the theory on trade and trade agreements.

Zero trade flows

Finally, the estimations presented thus far have ignored the issue of zero-trade flows. Trade was simply log-transformed, removing any zero trade flows from the regressions. Egger et al. (2011) have found that the resulting selection effects can bias the estimate of the effect of trade agreements downwards by as much as 35%. However, similar to the treatment of the binary trade agreement variable, it should be possible to control for this selection bias using a latent variable.[14]

Equation 6.14 is the multiplicative version of the gravity model used in the qualitative VAR model (equation 6.3).

$$X_{ij} = \exp\left(\phi_2\, RIA_{ij,t-1} + x_{ij,t}\, b_2 + c_{2_{ij}}\right)\, \zeta_{ij,t} \tag{6.14}$$

[13]Since the model is used to assess the impact of trade agreements, the initial categorization can only be based on ex ante differences in scope and depth of the agreement. Any indicator of its effectiveness should be left out.

[14]See for example Koop (2003) for the Bayesian estimation of a Tobit model using latent variables.

where ζ is drawn from a log-normal distribution $\mathcal{N}(0, \sigma_2)$.

By creating a new latent variable, $X^\star$, equation 6.14 can be log-transformed without losing the zero-trade flows. This latent trade variable could be seen as the desired trade given the present supply, demand and trade costs. If it is positive, trade will be equal to its exponent while if it is negative, trade is simply zero (Li, 1998).

$$
\begin{aligned}
X^\star_{ij} &= \phi_2 \, Y_{ij,t-1} + x_{ij,t} \, b_2 + c_{2_{ij}} + \varepsilon_{2_{ij,t}} && (6.15) \\
X_{ij,t} &= \begin{cases} 0 & \text{if } X^\star_{ij,t} \leq 0 \\ \exp(X^\star_{ij,t}) & \text{otherwise.} \end{cases} && (6.16)
\end{aligned}
$$

Combining this with the determinants of trade agreements and adding p lags, the qualitative gravity model becomes:

$$
\begin{aligned}
\begin{bmatrix} RIA^\star_{ij,t} \\ X^\star_{ij,t} \end{bmatrix} &= \sum_{k=1}^{p} \Phi^{(k)} \begin{bmatrix} RIA^\star_{ij,t-k} \\ X^\star_{ij,t-k} \end{bmatrix} + b \, x'_{ij,t} + c_{ij} + \varepsilon_{ij,t} && (6.17) \\
RIA_{ij,t} &= \begin{cases} 0 & \text{if } RIA^\star_{ij,t} \leq 0 \\ 1 & \text{otherwise} \end{cases} && (6.18) \\
X_{ij,t} &= \begin{cases} 0 & \text{if } X^\star_{ij,t} \leq 0 \\ \exp\left(X^\star_{ij,t}\right) & \text{otherwise.} \end{cases} && (6.19)
\end{aligned}
$$

Estimating the qualitative VAR model with the untruncated trade variable follows the approach outlined in section 6.3.2. To that end the state-space model is adjusted to draw values of both $X^\star$ and $RIA^\star$:

$$\begin{bmatrix} Y_{ij,t} \\ Y_{ij,t-1} \\ \vdots \\ Y_{ij,t-p+1} \end{bmatrix} = \begin{bmatrix} b x'_{ij,t} + c_{ij} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \Phi^{(1)} & \Phi^{(2)} & \cdots & \Phi^{(p)} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} Y_{ij,t-1} \\ Y_{ij,t-2} \\ \vdots \\ Y_{ij,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij,t} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{6.20}$$

$$\log\left(X_{ij,t}\right) = \begin{bmatrix} 0_{1\times m-1} & I_{m-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} Y_{ij,t} \\ Y_{ij,t-1} \\ \vdots \\ Y_{ij,t-p+1} \end{bmatrix} \quad \text{if } X_{ij,t} > 0 \tag{6.21}$$

$$0 = \begin{bmatrix} 0_{1\times m-1} & 0_{m-1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} Y_{ij,t} \\ Y_{ij,t-1} \\ \vdots \\ Y_{ij,t-p+1} \end{bmatrix} \quad \text{if } X_{ij,t} = 0 \tag{6.22}$$

6.7 Preliminary Conclusion

This chapter uses a qualitative VAR to bring together the literature explaining the causes of regional integration agreements with gravity equations in which its effects are measured. By taking the dynamic behavior of trade and RIAs into account, there is no need to look for the elusive instruments that affect trade but not agreements or vice versa. Furthermore, their endogenous relation can be identified without running afoul of any logical inconsistencies that pose a problem in cross-sectional studies.

Our preliminary findings confirm the usefulness of studying the behavior of trade and RIAs dynamically. An increase in trade motivates countries to sign integration agreements and an increase in the willingness to sign in turn raises trade. As could be expected, these effects take a long time to fully play out. Overall, the effect of trade agreements on trade and the average GDP are relatively small compared to what is typically found in the literature. The former increases quickly in the first 5 years, after which the growth slows down. Trade grows with 10% in the first year, 40% after 5 years and with about 50% after 10 years, while the average

GDP initially remains unaffected. After about 35 years, trade has risen 80% and average GDP with 10%. However, the model needs to be extended further if we want to compute reliable average treatment effects. The link between the qualitative VAR model and the theory on trade and trade agreements in particular needs to be strengthened before any final conclusions can be drawn.

References

Anderson, J.E. and van Wincoop, E. (2003) Gravity with gravitas: a solution to the border puzzle. *American Economic Review* 93(1):170–192.

Baier, S.L., Bergstrand, J.H. (2002) On the endogeneity of international trade flows and free trade agreements.

Baier, S.L., Bergstrand, J.H. (2004) Economic determinants of free trade agreements. *Journal of International Economics* 64(1):29–63.

Baier, S.L., Bergstrand, J.H. (2004) Do free trade agreements actually increases members' international trade?

Baier, S.L., Bergstrand, J.H. (2007) Do free trade agreements actually increases members' international trade? *Journal of International Economics* 71:72–95.

Baier, S.L., Bergstrand, J.H. (2009) Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of International Economics* 77:63–76.

Baldwin, R. and Taglioni, D. (2006) Gravity for dummies and dummies for gravity equations. *National Bureau of Economic Research working paper* 12516.

Dueker, M. (2005a) Dynamic forecasts of qualitative variables: A Qual VAR model of U.S. recessions. *Journal of Business & Economic Statistics* 23(1):96-104.

Dueker, M. (2005b) Kalman filtering with truncated normal state variables for

Bayesian estimation of macroeconomic models. Federal Reserve Bank of St. Louis working paper 2005-057.

Dueker, M. and Nelson C. R. (2005) Business cycle filtering of macroeconomic data via a latent business cycle index. *Macroeconomic Dynamics* 10:573–594.

Durbin, J. and Koopman, S. (2012) *Time series analysis by state space methods, 2nd edition.* Oxford University Press, Oxford.

Egger, P., Larch, M., Staub, K.E., and Winkelmann, R. (2011) The trade effects of endogenous preferential trade agreements. *American economic Journal: Economic Policy:* 113–143.

Egger, P. and Nigai, S. (2015) Structural gravity with dummies only. Centre for Economic Policy Research 10427.

Feenstra, R.C., Inklaar, R. and Timmer, M.P. (2013) The next generation of the Penn world table.

Frankel, J.A. (1997) *Regional trading blocs in the world economic system.* Peterson Institute Press, Washington, D.C.

Guimarães, P. and Portugal, P. (2009) A simple feasible alternative procedure to estimate models with high-dimensional fixed effects. *IZA Discussion paper* 3935.

Head, K., Mayer, T. and Ries, J. (2010) The erosion of colonial trade linkages after independence. *Journal of International Economics* 81(1):1–14.

Head, K. and Mayer, T. (2013) Gravity equations: toolkit, cookbook, workhorse. in Gopinath, G., Helpman, E. and Rogoff, K. (editors) *Handbook of international economics, Vol. 4*, Elsevier, Amsterdam.

Kim, C.J. and Nelson, C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approaches with applications.* MIT Press, Cambrigde.

Kohl, T., Brakman S. and Garretsen, G. (2014) Do trade agreements stimulate international trade differently? Evidence from 296 trade agreements.

Koop, G. (2003) *Bayesian econometrics.* John Wiley & sons, Chichester.

Koop, G. and Korobilis, D. (2009) Bayesian multivariate time series methods for empirical macroeconomics. Mumich Personal RePEc Archive paper 20125.

Krugman, P. (1995) The move toward free trade zones. In King P. (editor) *Internatinoal economics and international economic policy: a reader.* McGraw-Hill, New York, 163–198.

Levy, P. (1997) A political-economy analysis of free-trade agreements. *American Economice Review* 87(2):506–519.

Li, K. (1998) Bayesian inference in a simultaneous equation model with limited dependent variables. *Journal of Econometrics* 85: 387–400.

Magee, C. S. (2003) Endogenous preferential trade agreements: an empirical analysis. *Contributions to Economic Analysis & Policy* 2(1):1166–1217.

Márquez-Ramos, L., Martínez-Zarzoso, I., Suárez-Burguet, C. (2011) Determinants of deep integration: Examining socio-political factors. *Open Economic Review* 22(3):479–500.

Marshall, M.G., Gurr, T.R. and Jaggers, K. (2014) Polity IV project: political regime characteristics and transitions, 1800-2013. *Center for Systemic Peace*

Pesaran, M.H., Schuermann, T. and Weiner, S.T. (2004) Modeling regional interdependencies using a global error-correcting macroeconomic model. *Journal of business & economic statistics* 22(2):129–162.

Standaert, S. (2014) Divining the level of corruption: a Bayesian state-space approach. *Journal of Comparative Economics.*

Tinbergen, J. (1962) *Shaping the World Economy.* The Twentieth Century Fund, New York.

Wu, J. P. (2006) Measuring and explaining levels of regional economic integration. In De Lombaerde, P. (editor), *Assessment and Measurement of Regional Integration.* Routledge, New York, Ch. 9, pp. 162–179.

Appendices

6.A Estimating a Qual VAR

A. Drawing the paramater values

Let $z_{ij,t}^{(k)}$ be the vector of all exogenous and lagged endogenous explanatory variables in the k^{th} equation and m the number of endogenous variables, we can write the qualitative VAR model (6.3) as:

$$Y_{ij,t} = c_{ij} + Z_{ij,t}\beta + \varepsilon_{ij,t} \tag{6.23}$$

with

$$Z_{ij,t} = \begin{bmatrix} z_{ij,t}^{(1)} & 0 & \cdots & 0 \\ 0 & z_{ij,t}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_{ij,t}^{(m)} \end{bmatrix},$$

$Y_{ij,t} = [Y_{ij,t}^{(1)}, \cdots, Y_{ij,t}^{(m)}]'$, $\beta = [\beta^{(1)}, \cdots, \beta^{(m)}]'$ and $\varepsilon_{ij,t} = [\varepsilon_{ij,t}^{(1)}, \cdots, \varepsilon_{ij,t}^{(m)}]' \sim N(0, \Sigma)$.
The prior distribution of the parameters employed can be written as:

$$\beta \sim N(\underline{b}, \underline{V}_b) \tag{6.24}$$

$$\Sigma \sim iWish(\underline{S}, \underline{\nu}) \tag{6.25}$$

Using the the maximum likelihood estimates as a starting point, the Gibbs sampler progresses through the following conditional posterior probabilities (Koop and

Korobilis, 2009):

1. $\beta | c_{ij}, \Sigma, RIA^{\star}, Z \sim N\left(\bar{\beta}, \bar{V}_{\beta}\right)$

   with

   $$\bar{V}_{\beta} = \left(\underline{V}_{\beta}^{-1} + \sum_{ij,t} Z'_{ij,t} \Sigma^{-1} Z_{ij,t} \right)^{-1}$$

   and

   $$\bar{\beta} = \bar{V}_{\beta} \left(\underline{V}_{\beta}^{-1} \underline{\beta} + \sum_{ij,t} Z'_{ij,t} \Sigma^{-1} (Y_{ij,t} - c_{ij}) \right).$$

2. $c_{ij} | \beta, \Sigma, RIA^{\star}, Z \sim N(\bar{c}_{ij}, \bar{V}_{c_{ij}})$

   with

   $$\bar{c}_{ij} = \frac{\sum_{t=1}^{T_{ij}} (Y_{ij,t} - Z_{ij,t}\beta)}{T_{ij}}$$

   and $\bar{V}_{c_{ij}} = diag(\Sigma)/T_{ij}$. When controlling for sender-target fixed effects, T_{ij} is the number of observations covering the country-couple ij. With separate country fixed effects, it can be split up as$c_{ij} = c_i + c_j$. To estimate this step 2 is run twice: the first time grouping per sender and the second time per target (Guimarães and Portugal, 2009).

3. $\Sigma | \beta, RIA^{\star}, Z \sim iWish(S, T)$

   with

   $$S = \sum_{ij,t} (Y_{ij,t} - Z_{ij,t}\beta - c_{ij})(Y_{ij,t} - Z_{ij,t}\beta - c_{ij})'.$$

   and T the total number of observations. Σ is subsequently normalized such that the first diagonal element is one while preserving the correlation coefficients (Dueker, 2005a).

After a new value for the variance-covariance matrix is drawn, the new parameter values for Θ and Σ are used in the Kalman filter and Kalman smoother to draw new values for $RIA^{\star}$. The process is then repeated from step 1 until convergence has been achieved.

B. The adjusted Kalman filter and smoother

Unlike the estimation of the parameters Θ, the latent variable can be generated for each country-couple separately. Dropping the country indices and simplifying equations 6.7 and 6.8 reveals the familiar state-space model structure.

$$S_t = \mu_t + F S_{t-1} + \nu_t \tag{6.26}$$

$$X_t = H S_t \tag{6.27}$$

with $var(\varepsilon) = Q$.

Instead of having to estimate the entire model at once, the Kalman filter and smoother allows us to iteratively estimate and draw from the probability of the latent variable. Starting from $t = 0$, the Kalman filter iterates forward through time, computing the mean and variance of $RIA^{\star}$ at time t, conditional on all information up until that moment. After completing the Kalman filter, a standard simulation smoother algorithm can be used to draw values of $RIA^{\star}$ starting at the final observation and iterating backward. The end result is a new draw of the latent variable which contain all information in the dataset. These can subsequently be used to re-estimate the parameters of the qualitative VAR model. More information on the Kalman filter and simulation smoother can be found in Kim and Nelson (1999) and Durbin and Koopman (2012).

The difference with a normal Kalman filter is that the value of RIA has to be taken into account. The expected value and variance of the latent variable changes depending on whether the countries have signed an agreement or not. First, the distribution of S_t is predicted using the outcome of the previous iteration (Dueker, 2005b):

$$S_{t|t-1} = E(S_t|S_{t-1}) = \mu_t + F\, S_{t-1|t-1} + E(\nu_t|RIA_t) \tag{6.28}$$

$$P_{t|t-1} = var(S_t|S_{t-1}) = F\, P_{t-1|t-1}\, F' + var(\nu_t|RIA_t) \tag{6.29}$$

Let F_1 and μ_{t_1} be the first row of matrix F and the first element of the vector μ_t. If we define $\tau = \mu_{t_1} + F_1\, S_{t|t}$ and using ϕ and Φ to denote the normal pdf and cdf, the conditional distribution of ν_t can be written as:

$$E(\nu_t|RIA_t) = a = \begin{cases} -\frac{\phi(\tau)}{\Phi(-\tau)} & \text{if } RIA_t = 0 \\ \frac{\phi(\tau)}{1-\Phi(-\tau)} & \text{if } RIA_t = 1 \end{cases} \tag{6.30}$$

$$var(\nu_t|RIA_t) = \begin{cases} 1 - a^2 + \frac{\tau\,\phi(\tau)}{\Phi(-\tau)} & \text{if } RIA_t = 0 \\ 1 - a^2 - \frac{\tau\,\phi(\tau)}{1-\Phi(-\tau)} & \text{if } RIA_t = 1 \end{cases} \tag{6.31}$$

After prediction, $RIA^\star$ is subsequently updated using the information contained in the measurement equation. The difference between the two is called the Data Forecast Error, while the weight the new information receives, κ, is the Kalman gain.

$$DFE = X_t - H\, S_{t|t-1} \tag{6.32}$$

$$\kappa_t = P_{t+1|t}\, H'(H\, P_{t|t-1}\, H')^{-1} \tag{6.33}$$

$$S_{t|t} = S_{t|t-1} + \kappa_t\, DFE \tag{6.34}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t\, H\, P_{t|t-1} \tag{6.35}$$

After the Kalman filter has completed, a normal Kalman smoother can be used to compute the distribution of $RIA^\star$ using all available information, which can be drawn from using a truncated normal distribution (Dueker, 2005b).

6.B List of the regional integration agreements

Table 6.2: List of the regional integration agreements

Andean Community of Nations	Georgia - Azerbaijan
Arab Maghreb Union (AMU)	Georgia - Kazakhstan
Armenia - Kazakhstan	Georgia - Russian Federation
Armenia - Moldova	Georgia - Turkmenistan
Armenia - Russian Federation	Georgia - Ukraine
Armenia - Turkmenistan	Guatemala - Chinese Taipei
Armenia - Ukraine	Hong Kong, China - Chile
ASEAN - Australia - New Zealand	Hong Kong, China - New Zealand
ASEAN - China	Iceland - China
ASEAN - India	Iceland - Faroe Islands
ASEAN - Japan	India - Bhutan
ASEAN - Korea, Republic of	India - Japan
ASEAN Free Trade Area (AFTA)	India - Malaysia
Australia - Chile	India - Singapore
Australia - New Zealand (ANZCERTA)	India - Sri Lanka
Australia - Papua New Guinea (PATCRA)	Indian Ocean Commission (IOC)
Brunei Darussalam - Japan	Intergovernmental Authority on Development (IGAD)
Canada - Chile	Israel - Mexico
Canada - Colombia	Japan - Australia
Canada - Costa Rica	Japan - Indonesia
Canada - Israel	Japan - Malaysia
Canada - Jordan	Japan - Mexico
Canada - Panama	Japan - Peru
Canada - Peru	Japan - Philippines
Canada - Rep. of Korea	Japan - Singapore
Caribbean community CARICOM / CARIFORUM	Japan - Switzerland
Caribbean free trade association (CARIFTA)	Japan - Thailand
Caribbean single market and economy (CSME)	Japan - Viet Nam
Central European Free Trade Agreement (CEFTA)	Jordan - Singapore
Chile - China	Korea, Republic of - Australia
Chile - Colombia	Korea, Republic of - Chile
Chile - Costa Rica (Chile - Central America)	Korea, Republic of - India
Chile - El Salvador (Chile - Central America)	Korea, Republic of - Singapore
Chile - Guatemala (Chile - Central America)	Korea, Republic of - Turkey
Chile - Honduras (Chile - Central America)	Korea, Republic of - US
Chile - Japan	Kyrgyz Republic - Armenia
Chile - Malaysia	Kyrgyz Republic - Kazakhstan
Chile - Mexico	Kyrgyz Republic - Moldova
Chile - Nicaragua (Chile - Central America)	Kyrgyz Republic - Russian Federation
China - Costa Rica	Kyrgyz Republic - Ukraine
China - Hong Kong, China	Kyrgyz Republic - Uzbekistan
China - Macao, China	Latin America Free Trade Association LAFTA / LAIA
China - New Zealand	Malaysia - Australia
China - Singapore	Mano River Union (MRU)
Colombia - Mexico	Melanisian spearhead group (MSG) trade agreement
Colombia - Northern Triangle (El Salvador, Guatemala, Honduras)	Mexico - Central America
Common Economic Zone (CEZ)	Mexico - Uruguay
Common Market for Eastern and Southern Africa (COMESA)	New Zealand - Chinese Taipei

Common market of the South (MERCUSOR)
Commonwealth of Independent States (CIS)
Community of Sahel-Saharan States (CEN-SAD)
Costa Rica - Peru
Costa Rica - Singapore
Dominican Republic - Central America
Dominican Republic - Central America - United States Free Trade Agreement (CAFTA-DR)
East African Community (EAC)
Economic and Monetary Community of Central Africa (CEMAC)
Economic Community of Central African States (ECCAS)
Economic Community of the Great Lakes Countries (CEPGL)
Economic Community of West African States (ECOWAS)
EFTA
EFTA - Albania
EFTA - Bosnia and Herzegovina
EFTA - Canada
EFTA - Central America (Costa Rica and Panama)
EFTA - Chile
EFTA - Colombia
EFTA - Egypt
EFTA - Former Yugoslav Republic of Macedonia
EFTA - Hong Kong, China
EFTA - Israel
EFTA - Jordan
EFTA - Korea, Republic of
EFTA - Lebanon
EFTA - Mexico
EFTA - Montenegro
EFTA - Morocco
EFTA - Palestinian Authority
EFTA - Peru
EFTA - SACU
EFTA - Serbia
EFTA - Singapore
EFTA - Tunisia
EFTA - Turkey
EFTA - Ukraine
Egypt - Turkey
El Salvador- Honduras - Chinese Taipei
EU - Albania
EU - Algeria
EU - Bosnia and Herzegovina
EU - Cameroon
EU - CARIFORUM States EPA
EU - Central America
EU - Chile
EU - Colombia and Peru
EU - CÃ´te d'Ivoire
EU - Eastern and Southern Africa States Interim EPA
EU - Egypt
EU - Faroe Islands
New Zealand - Malaysia
New Zealand - Singapore
Nicaragua - Chinese Taipei
North American Free Trade Agreement (NAFTA)
Pacific Island Countries Trade Agreement (PICTA)
Pakistan - China
Pakistan - Malaysia
Pakistan - Sri Lanka
Pan-Arab Free Trade Area (PAFTA)
Panama - Chile
Panama - Chinese Taipei
Panama - Costa Rica (Panama - Central America)
Panama - El Salvador (Panama - Central America)
Panama - Guatemala (Panama - Central America)
Panama - Honduras (Panama - Central America)
Panama - Nicaragua (Panama - Central America)
Panama - Peru
Panama - Singapore
Peru - Chile
Peru - China
Peru - Korea, Republic of
Peru - Mexico
Peru - Singapore
Russian Federation - Azerbaijan
Russian Federation - Belarus
Russian Federation - Kazakhstan
Russian Federation - Republic of Moldova
Russian Federation - Serbia
Russian Federation - Tajikistan
Russian Federation - Turkmenistan
Russian Federation - Uzbekistan
Singapore - Australia
Singapore - Chinese Taipei
South Asian Free Trade Agreement (SAFTA)
Southern Africa Customs Union (SACU)
Southern African Development Community (SADC)
Switzerland - China
Thailand - Australia
Thailand - New Zealand
Trans-Pacific Strategic Economic Partnership
Treaty on a Free Trade Area between members of the Commonwealth of Independent States (CIS)
Turkey - Albania
Turkey - Bosnia and Herzegovina
Turkey - Chile
Turkey - Former Yugoslav Republic of Macedonia
Turkey - Georgia
Turkey - Israel
Turkey - Jordan
Turkey - Mauritius
Turkey - Montenegro
Turkey - Morocco

EU - Former Yugoslav Republic of Macedonia	Turkey - Palestinian Authority
EU - Georgia	Turkey - Serbia
EU - Iceland	Turkey - Syria
EU - Israel	Turkey - Tunisia
EU - Jordan	Ukraine - Azerbaijan
EU - Korea, Republic of	Ukraine - Belarus
EU - Lebanon	Ukraine - Former Yugoslav Republic of Macedonia
EU - Mexico	Ukraine - Kazakhstan
EU - Montenegro	Ukraine - Moldova
EU - Morocco	Ukraine - Montenegro
EU - Norway	Ukraine - Russian Federation
EU - Overseas Countries and Territories (OCT)	Ukraine - Tajikistan
EU - Palestinian Authority	Ukraine - Uzbekistan
EU - Papua New Guinea / Fiji	Ukraine -Turkmenistan
EU - Rep. of Moldova	US - Australia
EU - Serbia	US - Bahrain
EU - South Africa	US - Chile
EU - Switzerland - Liechtenstein	US - Colombia
EU - Syria	US - Israel
EU - Tunisia	US - Jordan
EU - Ukraine	US - Morocco
Eurasian Economic Union (EAEU)	US - Oman
European Union (EU)	US - Panama
Faroe Islands - Norway	US - Peru
Faroe Islands - Switzerland	US - Singapore
Georgia - Armenia	West African Economic and Monetary Union (UEMOA)

6.C Summary statistics

Table 6.3: Summary statistics

Variable	Source	Obs	Mean	Std. Dev.	Min	Max
RIA	WTO	854,959	0.046	0.209	0	1
Trade	DoTS	455,521	15.029	3.602	−27.457	27.2
GDP_{av}	pwt 8.0	485,112	10.344	1.473	4.925	16.277
GDP_{diff}	pwt 8.0	485,112	2.488	1.837	$1.40e-6$	11.231
DKL	pwt 8.0	480,457	1.663	1.186	$3.59e-6$	6.184
DROWKL	pwt 8.0	480,452	1.287	0.644	0.001	4.231
Distance	CEPII	854,959	8.733	0.772	4.107	9.892
Remote	CEPII	854,959	2.179	3.835	0	9.517
Landlocked	CEPII	854,959	0.276	0.447	0	1
Contiguity	CEPII	854,959	0.019	0.137	0	1
Common language	CEPII	816,181	0.170	0.376	0	1
Colony	CEPII	816,181	0.015	0.122	0	1
Common colony	CEPII	816,181	0.115	0.319	0	1
WTO	WTO	854,959	0.341	0.474	0	1
Pol. comp	Polity IV	516188	9.900	21.810	0	98
Durability	Polity IV	514757	27.089	31.865	0	204
Autocracy	Polity IV	516188	9.391	21.120	0	98

6.D Reduced parameter values of the full model - World

Table 6.4: Reduced parameter values of the full model - World

	RIA*	st.e.	X	st.e.	DKL	st.e.	GDP$_{diff}$	st.e.	GDP$_{av}$	st.e.
L1.RIA*	0.8751^{a}	(0.0781)	-0.0024	(0.0023)	-0.0046^{a}	(0.0019)	0.0006	(0.0004)	0.0002	(0.0005)
L2.RIA*	-0.0045	(0.0140)	0.0008	(0.0008)	-0.0003	(0.0004)	0.0001	(0.0002)	0.0001	(0.0002)
L3.RIA*	-0.0019	(0.0082)	0.0001	(0.0006)	0.0004	(0.0004)	-0.0000	(0.0002)	0.0003	(0.0003)
L1.X	0.0023	(0.0026)	0.8811^{a}	(0.0008)	-0.0005^{a}	(0.0002)	0.0012^{a}	(0.0001)	0.0007^{a}	(0.0002)
L2.X	0.0005	(0.0006)	0.0058^{a}	(0.0005)	-0.0001	(0.0001)	0.0005^{a}	(0.0001)	-0.0012^{a}	(0.0002)
L3.X	0.0003	(0.0003)	0.0031^{a}	(0.0003)	-0.0001	(0.0001)	0.0002^{a}	(0.0001)	-0.0004^{a}	(0.0001)
L1.DKL	-0.0454^{a}	(0.0197)	-0.0125^{a}	(0.0036)	0.9783^{a}	(0.0023)	0.0008	(0.0009)	-0.0116^{a}	(0.0015)
L2.DKL	-0.0105^{a}	(0.0044)	-0.0038^{b}	(0.0022)	0.0043^{a}	(0.0005)	0.0003	(0.0006)	0.0048^{a}	(0.0011)
L3.DKL	-0.0046^{a}	(0.0017)	-0.0016	(0.0015)	-0.0023^{a}	(0.0003)	0.0001	(0.0004)	0.0055^{a}	(0.0009)
L1.GDP$_{diff}$	-0.0056	(0.0060)	-0.0357^{a}	(0.0024)	0.0027^{a}	(0.0004)	0.9715^{a}	(0.0006)	-0.0013	(0.0009)
L2.GDP$_{diff}$	-0.0014	(0.0017)	-0.0086^{a}	(0.0016)	-0.0002	(0.0003)	0.0123^{a}	(0.0004)	-0.0027^{a}	(0.0008)
L3.GDP$_{diff}$	-0.0006	(0.0009)	-0.0037^{a}	(0.0011)	-0.0003^{b}	(0.0002)	0.0044^{a}	(0.0003)	-0.0036^{a}	(0.0005)
L1.GDP$_{av}$	0.0018	(0.0041)	0.0131^{a}	(0.0015)	-0.0006	(0.0005)	0.0072^{a}	(0.0004)	0.9586^{a}	(0.0008)
L2.GDP$_{av}$	0.0005	(0.0011)	0.0032^{a}	(0.0008)	-0.0001	(0.0002)	0.0017^{a}	(0.0002)	0.0088^{a}	(0.0005)
L3.GDP$_{av}$	0.0002	(0.0005)	0.0015^{a}	(0.0004)	-0.0000	(0.0001)	0.0007^{a}	(0.0001)	0.0027^{a}	(0.0003)
Distance	-0.5075^{a}	(0.1534)	-0.0926^{a}	(0.0049)	-0.0099^{a}	(0.0028)	0.0101^{a}	(0.0008)	0.0139^{a}	(0.0007)
DROWKL	1.7331^{a}	(0.2874)	0.1956^{a}	(0.0187)	0.0455^{a}	(0.0070)	-0.0564^{a}	(0.0027)	0.0239^{a}	(0.0019)
Landlocked	-0.3718^{b}	(0.2343)	-0.0053	(0.0189)	-0.0013	(0.0030)	-0.0167^{a}	(0.0023)	0.0121^{a}	(0.0012)
Remote	0.0338^{a}	(0.0364)	0.0089^{a}	(0.0012)	-0.0010^{a}	(0.0003)	-0.0003^{c}	(0.0002)	0.0014^{a}	(0.0001)
Contiguous	-0.0666	(0.2311)	0.1291^{a}	(0.0130)	-0.0073^{a}	(0.0024)	-0.0020	(0.0017)	0.0105^{a}	(0.0011)
Colony	0.0020	(0.0800)	0.1448^{a}	(0.0181)	0.0044^{b}	(0.0023)	0.0070^{a}	(0.0023)	-0.0027^{c}	(0.0015)
Common Colony	0.4354^{b}	(0.1199)	0.1360^{a}	(0.0121)	-0.0001	(0.0022)	-0.0247^{a}	(0.0018)	0.0043^{a}	(0.0010)
Language	0.1287^{a}	(0.1223)	0.0551^{a}	(0.0086)	0.0017	(0.0011)	-0.0040^{a}	(0.0012)	-0.0017^{a}	(0.0006)
WTO	1.0482^{a}	(0.4330)	0.1876^{a}	(0.0103)	0.0045^{c}	(0.0024)	-0.0292^{a}	(0.0015)	0.0245^{a}	(0.0008)
Autocracy	0.0227^{a}	(0.0150)	-0.0024^{c}	(0.0013)	0.0009^{a}	(0.0002)	0.0011^{a}	(0.0002)	-0.0012^{a}	(0.0001)
Pol. comp	-0.0230^{a}	(0.0138)	0.0016	(0.0012)	-0.0007^{a}	(0.0002)	-0.0006^{a}	(0.0002)	0.0008^{a}	(0.0001)
Durability	0.0077^{a}	(0.0030)	0.0008^{a}	(0.0001)	0.0001^{a}	(0.0000)	-0.0001^{a}	(0.0000)	0.0001^{a}	(0.0000)
	1669741		1669741		1669741		1669741		1669741	
	sender & target		sender & target		sender & target		sender & target		sender & target	

Reduced parameter estimates of the full, worldwide qualitative VAR model with three lags.
Standard errors between brackets. ^{a}, ^{b} and ^{c} indicate significance at the 1%, 5% and 10% level.

7 | Conclusion

Except for the second chapter, the bulk of this thesis is concerned with the creation of tools that will be used in later analyses. For example, the penultimate chapter introduces the qualitative VAR methodology to the literature on the effect of trade agreements on trade flows, but this work is far from completed. The third chapter in particular only discusses the construction of the Bayesian index of corruption, but does not use it to answer a specific research question. Similarly, we have barely scratched the surface of the uses of the indexes of actual economic integration (chapter 4) and historical trade integration (chapter 5). For this reason, instead of simply repeating the conclusions of each chapter, this final chapter discusses ongoing and planned future research that is based on the work presented in the preceding chapters.

The Bayesian Corruption Index

A significant portion of the third chapter is devoted to the comparison of the newly constructed Bayesian Corruption Index with the corruption estimate of the Worldwide Governance Indicators. In this chapter, we show that the BCI has an increased coverage, makes estimates with greater certainty and enables us to completely take its uncertainty into account in any subsequent analysis. However, to be useful from a research and policy perspective, the construction of the BCI still requires quite some work. To start with, the corruption index should be updated as its current estimates end in 2013. Secondly, the chapter lacks many of the robustness checks that were performed on the WGI. For example, it remains to be checked to what extent the index changes when each indicator receives the same weight (cf. Kaufmann et al., 2007a). Most importantly, the interpretability and ease of use should be vastly improved, if the BCI has any hopes of becoming an accepted measure of corruption.

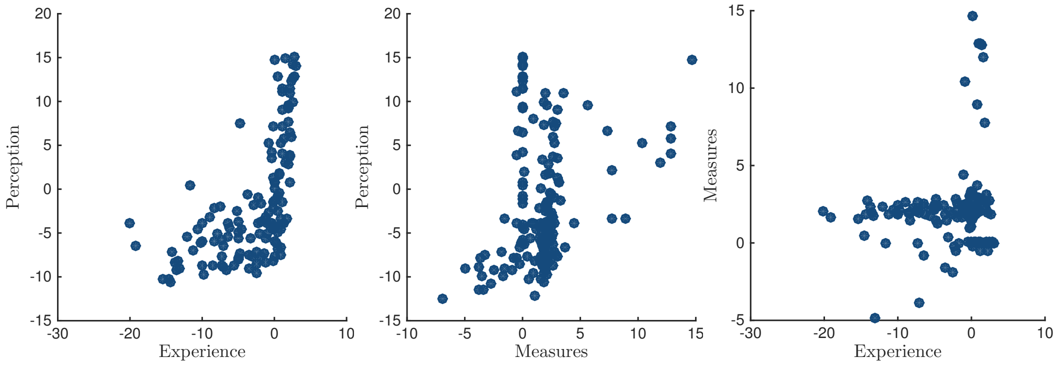
The specification of the state-space model could also be further refined. For example, as it is defined in chapter 3, the model ignores the exact nature of the underlying variables. Particularly in the case of limited dependent variables, the assumptions of the state-space model might be incompatible with the nature of the observed indicators. Following Høyland et al. (2012), this issue can be solved by imposing a Poisson distribution (in stead of a normal distribution) on some of the underlying indicators.

In terms of the applications of the BCI index, one possible research idea would be to look at the relationship between the perception of corruption, the actual experience with corruption and the presence of anti-corruption measures (cf. Roca, 2011). To that end, the methodology of the BCI (an index of corruption perception) was used to construct indicators that capture anti-corruption *measures* and *experience* with corruption.[1] Like the BCI, a high value of the *experience* index means that there has been little experience with corruption. In contrast, a high value of the *measures* index means that many anti-corruption measures are in effect.

Clearly, the relationship between these three indexes is not straightforward as each will in some way affect the other two. Limiting this preliminary examination to a simple scatterplot, figure 7.1 reveals a number of striking patterns. Without implying any causal relationship, it seems that when corruption is perceived to be rampant, large changes in the experience with corruption do not change the perceived level much (panel a). However, if the country is perceived to be less corrupt, a small increase in experience can be strongly detrimental to the perceived level of corruption. In addition, while an increase in the anti-corruption measures has a beneficial effect on the perceived level of corruption (panel b), its correlation with

[1]The indicators of experience with corruption and anti-corruption measures come from the same sources as the perception indicators listed in chapter 3, but were not used in the construction of the BCI.

the experience with corruption is much lower (panel c). A full analysis of the interplay between these three indices could potentially shed more light on the behavior of corruption over time, the usefulness of anti-corruption measures, as well as the value of indexes of corruption perception.



(a) Experience - Perception **(b)** Measures - Perception **(c)** Experience - Measures

Figure 7.1: Scatter plots of the index values of the perception of corruption, experience with corruption and anti-corruption measures in 2010

The historical trade index

The analysis of the distance puzzle in the worldwide trade network in the fifth chapter is only the first step in a broader research project that is centered around the historical trade index. Benjamin Vandermarliere, Stijn Ronsse and I are currently using the hierarchical stochastic block model method described in Peixoto (2015) to reveal the structure of the historical trade network over time.[2] A stochastic block model describes the underlying probability model of a network. Simply put, it divides the different countries (nodes) into blocks based on the similarity of their behavior in the network (i.e. the countries that they form edges to). It then describes for each block what the probability is that a member will trade with members of its own block, as well as each of the other blocks. Hierarchical block models also look for higher-level structures in the network: i.e. a group of blocks that behave

[2]The construction of this network and some simple network statistics are described in the appendix C of chapter 5.

similarly. The method of Peixoto is particularly interesting in that it uses Bayesian selection techniques to look for the most parsimonious model that accurately describes the network.

As an example, figure 7.2 plots the structure of the trade network in the period 1880-1914. In this graph, countries are represented by the dots on a circle. The bigger the dot, the larger its weighted indegree (the sum of all incoming edges). The nodes are sorted and colored according their membership to the lowest level block, as are their outgoing edges (indicating countries they are integrated into). The higher level groupings are represented by a thin red line connected to the center of the circle. To aid the interpretation of the graph, the groupings are also plotted on a world map.

Among other things, figure 7.2 shows that the world trade pattern over the period 1880-1914 corresponds to a core-periphery model. The core is defined as a group of countries that have mostly formed intra-block edges (to other countries in the core), while remaining relatively unconnected to the rest of the network: the periphery. The periphery in turn is strongly connected to the core, but its intra-block edges are sparse. The highest level grouping identifies the purple group (e.g. Great Britain and the USA) as the core and the rest of the world as the periphery. The periphery can be further subdivided into three groups. The yellow group is firmly in the periphery (e.g. Afghanistan and Peru), while the gray (e.g. Italy and Russia) and green (e.g. China and India) groups lie somewhere in between. The latter two have more intra-groups edges as well as a higher chance of receiving an edge from the core.

We are currently in the process of analyzing the world trade network on a yearly basis and tracking the evolution of the different groups from 1880 to 2010. In addition to revealing the changing structure of the network over time, we want to gauge the effect of strong shocks (e.g. the World Wars, oil crises, the great depression, etc.) on the network, paying particular attention to possible anticipatory changes.

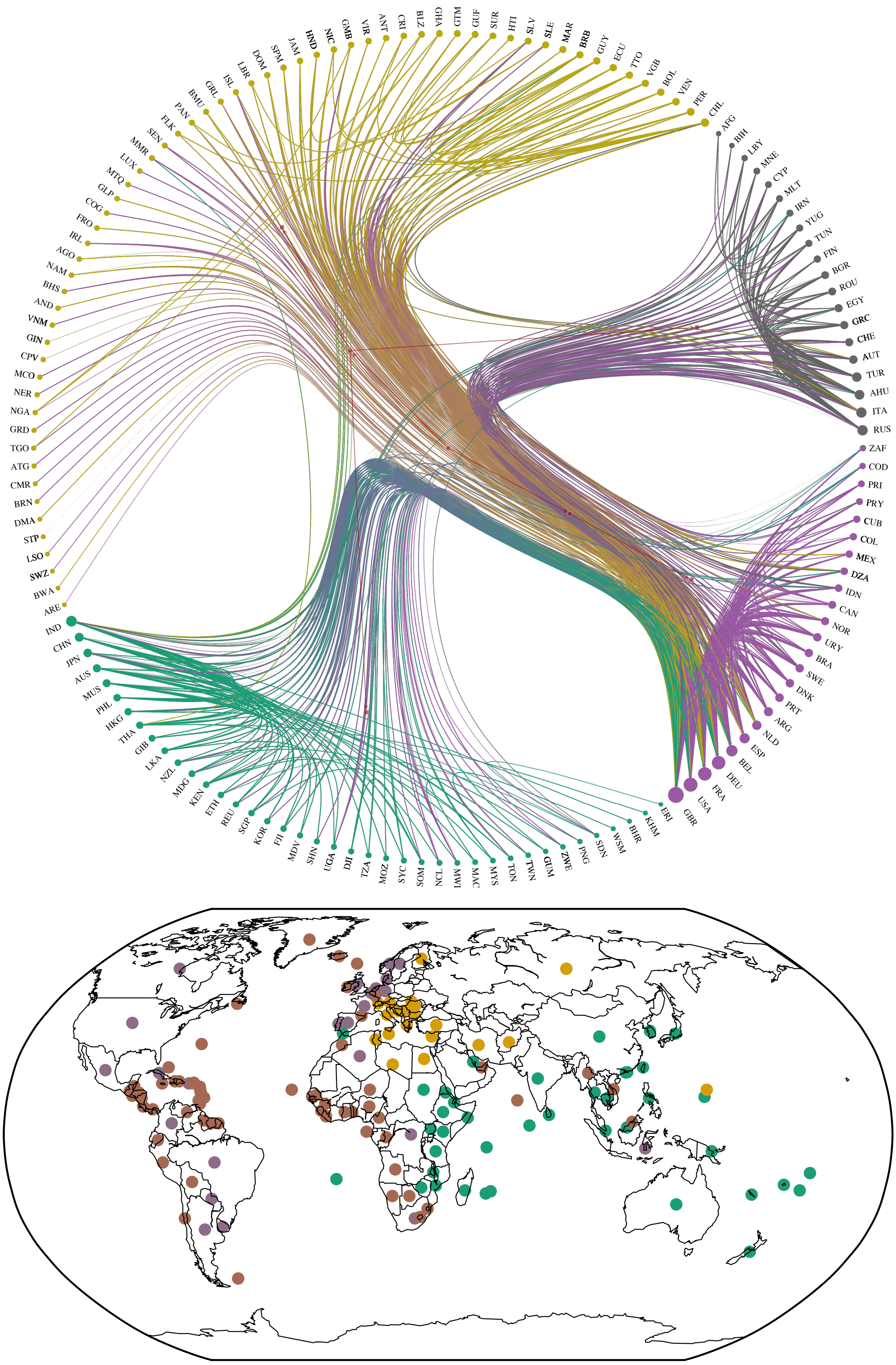


Figure 7.2: Structure of the historical trade network in the period 1880-1914
Nodes are sorted and colored according to the lowest level grouping of the block model, as are their outgoing edges. The bigger the node, the larger the weighted indegree. The higher level groupings are represented by a thin red line connected to the center of the circle.

State-space models

Needless to say, the idea of using a state-space model to create 'mashup' indicators can be applied to more than just integration and corruption.[3] The only requirement for this model to have a strong added advantage over more simple methods of aggregation (e.g. a principal components analysis) is that there is some level of persistence in the state variable. If this is combined with a complex pattern of missing variables, the advantage of using a state-space model rises even further.

One application Ilse Ruyssen, Glenn Rayp and I are currently working on is the creation of an index that captures the openness of migration policy. Persistence is likely to be high, as a change in policy requires some type of governmental decision. We have collected information on more than 250 indicators of the restrictiveness of migration policy, but the overabundance of missing values makes more simple aggregation methods impossible to use. However, as we explained in chapter 3, the state-space model can easily solve this problem of non-overlapping missing values. In order to correctly include some of the indicators, the model had to be adjusted to deal with binary indicators as well as indicators that can only be compared over time (but not over countries).

Other 'mashup' indexes we are currently working on are an index of the level of social protection on the labour market and an index of the level of competitiveness of the economy.

The qualitative vector autoregression model

As the extensions section in the previous chapter (section 6.6) makes clear, the work on the qualitative VAR is far from finished. Besides to the changes listed in this section, there are a number of other additions worth following up on. The first pertains the different theories on the interplay between corruption and integration highlighted in the first chapter. As stated in the introduction of this thesis, our

[3]'They've done studies, you know. 60% of the time, it works every time.' (Brian Fontana, Anchorman, 2004)

initial motivation for looking at the qualitative VAR was as a way to correct for the endogenous relation between corruption and trade agreements.

To get a first glimpse of the results, I reran the full model regression for the entire world from section 6.5.2 with average corruption (as measured by the BCI) as an additional endogenous variable. The average treatment effect of a trade agreement on the average BCI is positive, meaning that the level of corruption decreases after an agreement is signed. Over time, this positive effect slowly increases, but the confidence bounds indicate that it is borderline (in)significant. While it is impossible to gauge the intentions of policy makers in closing the agreements, these results do indicate that the rent-destruction and rent-shielding effects dominate in trade agreements worldwide.

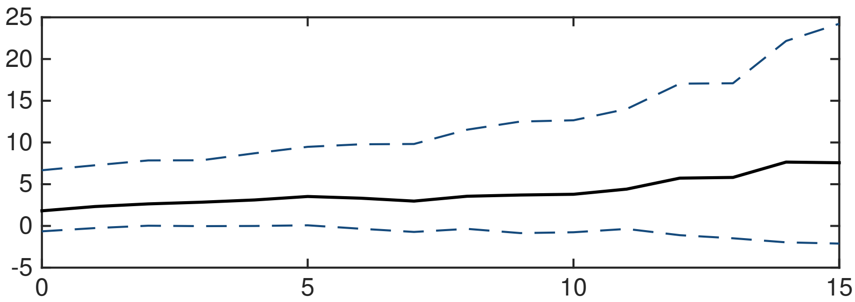


Figure 7.3: Average treatment effect of RIAs on corruption - World

A second addition to this model would be include the index of Actual Economic Integration of chapter 4 to study the effects of trade agreements on economic integration as a whole.

In conclusion, the chapters presented in this thesis have laid the groundwork for a large number of future research projects.

References

Høyland, B., Moene, K. and Willumsen, F. (2012) The tyranny of international index rankings. *Journal of Development Economics* 97(1):1–14.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2007a) The worldwide governance indicators: Answering the critics. World Bank Policy Research Paper 4194.

Peixoto, T.P. (2015) Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups. *Physical Review X* 5, 011033.

Roca, T. (2011) Measuring corruption: perception surveys or victimization surveys? Université Montesquieu-Bordeaux IV working paper DT/167/2011.